TEORIA E MÉTODOS DE MEDIDA EM CIÊNCIAS DO COMPORTAMENTO

LUIZ PASQUALI
(Organizador)

Educação e do Desporto

SEDIAE/INEP
Secretaria de Avaliação
e Informação Educacional
Instituto Nacional de Estudos e
Pesquisas Educacionais

LABPAM
Laboratório de Pesquisa
em Avaliação e Medida /
Instituto de Psicologia - UnB

TEORIA E MÉTODOS DE MEDIDA EM CIÊNCIAS DO COMPORTAMENTO

TEORIA E MÉTODOS DE MEDIDA EM CIÊNCIAS DO COMPORTAMENTO

LUIZ PASQUALI
(Organizador)

AGRADECIMENTOS

Muitas pessoas são responsáveis pelo final exitoso deste trabalho. Contudo, devo especiais agradecimentos à colaboração substancial dada pela equipe do Laboratório de Pesquisa em Avaliação e Medida do Instituto de Psicologia da Universidade de Brasília, sem cujo apoio este trabalho teria sido várias vezes mais dificultado. Notadamente, devo agradecer a Alexandre Damanico da Cunha, Cláudia Cristina Fukuda, Edilson Marcolino dos Santos, Heila Magali da Silva Veiga, Inaê de Figueiredo Cerqueira, Ivânia Ghesti e, em especial, a Maria Mazzarello Azevedo, vice-coordenadora do LÀBPAM.

O Organizador

SUMÁRIO

2ª PARTE — MÉTODOS E TÉCNICAS DE MEDIDA
EM CIÊNCIAS DO COMPORTAMENTO

PREFÁCIO

A importância da avaliação na área educacional é um consenso contemporâneo. O tema, até pouco tempo restrito à aprendizagem e desempenho dos alunos, vem ganhando a necessária amplitude. Estende-se objetivamente a projetos, programas, professores, currículos, instituições e sistemas, numa abordagem que destaca a interdependência de todos esses fatores.

Esta visão ampla e integrada é fundamental para responder à complexidade e totalidade do ato educativo e traduz o processo de consolidação da democracia brasileira. Ela expressa, sobretudo, a pressão social de cidadãos mais conscientes e mais exigentes quanto ao padrão dos serviços públicos que lhes são devidos.

Desde o início dos anos 90, o INEP tem sido um dos principais atores da implementação do Sistema de Avaliação de Ensino Básico (SAEB). Responsável pela avaliação do desempenho do sistema educacional, o SAEB atua em três eixos: rendimento do aluno, perfil e prática docente e gestão do ensino. A disseminação dos resultados obtidos com os três ciclos de aplicação do SAEB indica que o sistema vem se constituindo um marco importante na construção de uma cultura de avaliação já presente nas redes de ensino estaduais e municipais. Indica também que o SAEB vem cumprindo seu papel enquanto instrumento de suporte a formulação de estratégias e políticas educacionais que buscam elevar a qualidade e eqüidade da educação brasileira.

Um olhar crítico nessas experiências confirma a avaliação como um pressuposto ético e uma política pública necessária. Esse mesmo olhar crítico mostra, porém, que, para explorar todo o seu potencial e alcançar suas finalidades, ainda há muito a ser construído nas dimensões política, social e técnica da avaliação.

Este livro, que o INEP apresenta aos educadores — atuais e em formação — e à sociedade em geral, traz a contribuição de importantes especialistas em trabalhos que abordam os fundamentos, métodos e técnicas de medidas em ciências do comportamento. As questões aqui tratadas certamente agregarão novos elementos aos debates e práticas de avaliação que crescem em todo o País, sinalizando o compromisso com os alunos, com a escola, com a educação transformadora e com a construção de um país justo e igualitário.

Maria Helena Guimarães de Castro

APRESENTAÇÃO

A área da medida e da avaliação em ciências humanas e sociais, no Brasil, está quase totalmente carente de publicações nacionais. Os poucos textos existentes no mercado representam, em sua maioria, traduções de obras estrangeiras de anos há muito tempo idos. Como esta área vem mostrando um grande desenvolvimento nos últimos anos, essas traduções se ressentem de atualidade, deixando o pesquisador e profissional nacionais numa situação extremamente difícil de se aprofundar e atualizar nos fundamentos e técnicas da medida nas áreas de estudo do comportamento. O presente livro pretende ser o primeiro de uma série de obras nesta área, visando superar a lacuna mencionada. Este primeiro trabalho teve como intenção fundamental apresentar o *status* e os avanços da tecnologia da medida nos vários ramos da Psicologia, bem como cobrir a mais variada gama de enfoques que neles existem. O intento, creio, foi conseguido em sua grande parte. Apesar de especialistas em algumas áreas desta temática, após contatados, terem declinado de participar desta aventura, grande parte da temática da medida em Psicologia foi abordada. As áreas não abordadas neste livro dizem respeito, sobretudo, à medida piagetiana, à avaliação dita qualitativa e à análise experimental do comportamento. Evidentemente, outras muitas temáticas poderiam ser consideradas para entrar num livro deste teor, sobretudo no contexto da 2ª Parte, onde se trata mais de técnicas diferenciadas de avaliação do que dos fundamentos da medida, situação que poderá ser sanada em futuras edições, se for o caso.

O livro se divide em duas partes. A 1ª Parte — Fundamentos da Medida em Ciências do Comportamento — trata, obviamente, de aspectos epistemológicos e da fundamentação do enfoque quantitativista em ciências do comportamento, mais especificamente em Psicologia. Questões epistemológicas e estatísticas são, conseqüentemente, o tema desta parte. A teoria da medida é discutida no capítulo 1. Em seguida são tratados tipos diferentes de enfoques epistemológicos da medida em Psicologia, a saber, a medida psicofisica e a medida psicométrica. A

medida psicofisica é tratada no capítulo 2, enquanto os capítulos 3, 4, 5 e 7 tratam da medida psicométrica clássica e moderna e das escalas de atitude. O capítulo 6, enfim, anota problemas com o uso da Estatística em ciências do comportamento.

A 2ª Parte — Métodos e Técnicas de Medida em Ciências do Comportamento — trata do problema e das técnicas de mensuração ou de avaliação em áreas específicas da Psicologia. Assim, a medida na área da inteligência é tratada nos capítulos 8, 9 e 11 ; a observação do comportamento, no capítulo 10; a análise de conteúdo, no capítulo 12; as técnicas projetivas, no capítulo 13; a medida das atitudes, no capítulo 14; o capítulo 15 trata da técnica de *survey,* e o capítulo 16, da avaliação cognitivo-comportamental.

Além de visar instrumentar pesquisadores e profissionais das ciências do comportamento, este livro tem por objetivo iniciar uma discussão mais técnica, menos emocional, sobre a medida nestas ciências no País.

Brasília, setembro de 1996

Luiz Pasquali

OS AUTORES: PERFIL ACADÊMICO-PROFISSIONAL

Alfredo Oscar López Alonso

Licenciado em Psicologia e doutor em Filosofia e Letras, com orientação psicológica, pela Universidade de Buenos Aires. Professor titular, em nível de graduação e pós-graduacão, da Universidade de Buenos Aires, da Universidade de Belgrano, da Universidade de Salvador e da Universidade Nacional do Prata, na República Argentina. Sua especialização principal, tanto em docência como na pesquisa, tem sido em temas de medição em Psicologia, Psicometria, Metodologia da Pesquisa Psicològica e em Psicologia Cognitiva, inteligência humana, processos inferenciais e organização lógica do pensamento. É pesquisador principal do Conselho Nacional de Investigações Científicas e Técnicas da República Argentina (CONICET) e diretor do Instituto de Investigaciones Psicológicas da Universidad del Salvador, em Buenos Aires, Argentina, onde leva a termo seus trabalhos de pesquisa, como o que aqui se publica. É autor de mais de 60 trabalhos e publicações nas temáticas acima evocadas.

Cecília Guarnieri Batista

Psicóloga pela Pontifícia Universidade Católica de Campinas (PUC-CAMP) em 1972, concluiu o mestrado (1978) e o doutorado (1989) em Psicologia pelo Departamento de Psicologia Experimental do Instituto de Psicologia da Universidade de Sao Paulo (USP). Lecionou na UNIP (São Paulo) e na PUC-CAMP, notadamente Psicologia do Desenvolvimento. Participou na elaboração de projetos relacionados à prevenção de deficiências e à implantação do Programa de Integração e Desenvolvimento da Criança e do Adolescente (Prodecad), atualmente subordinado à Pró-Reitoria de Extensão e Assuntos Comunitários da UNICAMP. Atuou como psicóloga escolar do Prodecad, especialmente com os filhos de funcionários de baixa renda de universidade que cursavam da 1ª à 4ª série do primeiro grau (Programa de Apoio à Escolaridade). Presentemente, como docente de Educação Especial e Reabilitação no Centro de Estudos e

Pesquisas Prof. Dr. Gabriel Porto (CEPRE), da Faculdade de Ciências Médicas da UNICAMP, trabalha na área de deficiência visual, auditiva e múltipla. Seu trabalho de pesquisa tem envolvido o emprego e a reflexão teòrica sobre a Metodologia Observacional, o estudo da interação mãe-filho e o desenvolvimento cognitivo e moral infantil.

Edson A. de Souza Filho

Doutor em Psicologia Social pela École des Hautes Études en Sciences Sociales, em Paris, foi professor adjunto do Instituto de Psicologia da Universidade de Brasília (UnB) até bem pouco tempo. Tem ensinado e pesquisado em representações sociais, análise de conteúdo **e** do discurso, grupos e influência social. Atualmente é professor da Universidade Estadual do Norte Fluminense, em Campos, RJ.

Eleonora Vivas

Mestre e doutora (M.Phil. e Ph.D.) em Psicologia pela Universidade de Londres. Professora de pós-graduacão em Psicologia da Universidad Simon Bolívar, em Caracas, Venezuela. Trabalha nas áreas de avaliação comportamental, teoria e fundamentos do enfoque cognitivo-comportamental. Pesquisadora na área da motivação e prevenção de problemas de linguagem e leitura — psicologia da saúde.

Eunice Maria Lima Sorlano de Alencar

Licenciada em Psicologia pela Universidade Federal de Minas Gerais (UFMG), com mestrado e doutorado em Psicologia pela Purdue University e *post-doctoral scholar* no Gifted Education Resource Institute (Lafayette, Indiana, EUA). É professora titular da Universidade de Brasília (UnB), onde se dedica principalmente ao estudo da Criatividade, e, também, membro do Conselho Regional de Psicologia e do Conselho Técnico da Associação Brasileira para o Superdotado, tendo já presidido essa Associação. Além de autora de diversos livros, é, ainda, redatora da revista *Psicologia: Teoria e Pesquisa,* editada pelo Instituto de Psicologia **da** UnB.

Hartmut Günther

Nascido na Alemanha, estudou na Universitãt Hamburg **e** na Universitát Marburg. Graduou-se em Psicologia pelo Albion College, em Michigan, EUA. Tendo realizado o mestrado em Psicologia Experimental na Western Michigan University e o doutorado em Psicologia Social na University of California at Davis, é atualmente pesquisador e professor titular da Universidade de Brasília (UnB). Tem vários artigos publicados, tanto no Brasil como no exterior.

Jorge de Souza

Bacharel em Ciências Estatísticas pela Escola Nacional de Ciências Estatísticas, mestre em Ciências da Engenharia de Produção e doutor em Estatística pela Universidade Federal do Rio de Janeiro (UFRJ). É professor titular da Universidade de Brasília (UnB), com vários trabalhos publicados na área de Estatística, estando atualmente, como professor visitante, no Departamento de Economia da Universidade Federal de Pernambuco (UFPE). [Fax (081) 271-8360]

Jurema Alcides Cunha

Livre-docente em Psicologia Clínica e doutora em Filosofia, com experiência de ensino e pesquisa no Brasil e nos Estados Unidos. É professora dos cursos de Pós-Graduação em Psicologia da Pontifícia Universidade Católica do Rio Grande do Sui (PUC-RS) e já é conhecida por suas publicações, em especial pelo livro *Psicodiagnóstico-R,* adotado em muitas universidades brasileiras, e pelo Teste *das Fábulas — Forma Verbal e Pictórica.*

Leandro da Silva Almeida

Nascido em Portugal, é licenciado em Psicologia (1975-1980) e doutor em Psicologia (1987), na especialidade de Psicologia da Educação, pela Faculdade de Psicologia e de Ciências da Educação da Universidade do Porto. Atualmente é professor da Universidade de Braga e, ainda, membro da Comissão Científica da Unidade de Ciências da Educação, do Conselho Científico do CEFOPE, do Projeto de Investigação PIDACS (Universidade de Aveiro), do National Council on Measurement in Education (NCME), da International School Psychology Association (ISPA), entre outras. São de sua autoria os livros *Teoria da Inteligência* e *O Raciocínio Diferencial dos Jovens: Avaliação, Desenvolvimento e Diferenciação,* além de vários artigos sobre testes e educação.

Luiz Pasquali

Licenciado em Filosofia, Teologia, Pedagogia e Psicologia. Mestre e doutor em Psicologia pela Université Catholique de Louvain, Bélgica (1970). É professor titular da Universidade de Brasília (UnB), onde é responsável pelas disciplinas de Psicometria (Uso dos Testes), Métodos e Medidas em Psicologia e Análise Fatorial em Psicologia (Pós-Graduação), coordenando também o Curso de Especialização em Avaliação e Medida da UnB. É coordenador do Laboratório de Pesquisa em Avaliação e Medida do Instituto de Psicologia da Universidade de Brasília. Tem cerca de 40 trabalhos sobre instrumental psicológico.

Maria Alice Magalhães D'Amorim

Licenciada em Filosofia e em Psicologia pela Pontifícia Universidade Católica do Rio de Janeiro (PUC-RJ), tendo completado sua formação em Psicologia Aplicada na Universidade de Louvain, Bélgica. Mestre pela State University of lowa, EUA, fez o doutorado em Psicologia na Universidade de Louvain, pós-doutorado em Administração de Programas de Saúde Mental na University of North Carolina, EUA, **e** na Unidade de Psicologia do Trabalho e das Organizações da Universidade de Louvain — Louvain-la-Neuve, Bélgica. Lecionou na Pontifícia Universidade Católica do Rio de Janeiro (PUC-RJ), na Universidade de Québec em Trois Rivières, Canadá, na Universidade Federal da Paraíba (UFPB), na Universidade de Brasília (UnB) **e** na Universidade Federal do Rio Grande so Sul (UFRGS); encontra-se atualmente na Universidade Gama Filho, na cidade do Rio de Janeiro.

Maria Ângela Guimarães Feitosa

Formou-se em Psicologia pela Pontifícia Universidade Católica de São Paulo (PUC-SP), em 1971. Fez o mestrado em 1977 e o doutorado em Psicobiologia em 1983 na University of Michigan, USA. Foi diretora do Instituto de Psicologia da Universidade de Brasília (UnB) de 1987 a 1989 e é atualmente editora da revista *Psicologia: Teoria e Pesquisa.* Orientou três teses de mestrado em Psicobiologia e possui seis publicações na área nos últimos cinco anos. Ensina e pesquisa na área de Psicobiologia, especializando-se em processos sensoriais, particularmente audição.

Maria Lúcia Tiellet Nunes

Doutora pela Universidade Livre de Berlim, onde realizou também formação terapêutica. É professora nos cursos de graduação e pós-graduacão em Psicologia da Pontifícia Universidade Católica do Rio Grande do Sul (PUC-RS) e no curso de graduação em Psicologia da Universidade Federal do Rio Grande so Sul (UFRGS). Colaborou no livro *Psicodiagnóstico-R* **e** é co-autora de Tesfe *das Fábulas — Forma Verbal e Pictórica.*

1ª PARTE

FUNDAMENTOS DA MEDIDA EM CIENCIAS DO COMPORTAMENTO

CAPÍTULO 1

A TEORIA DA MEDIDA

Luiz Pasquali
Instituto de Psicologia
Universidade de Brasilia

A medida em ciências do comportamento, notadamente na Psicologia, deveria ser chamada puramente de psicometria, similarmente ao que ocorre em ciências afins a ela, onde se fala de sociometria, econometria, politicometria, etc. Psicometria, contudo, tem sido abusivamente utilizada dentro de um contexto muito restrito, referindo-se atestes psicológicos e escalas psicométricas. De qualquer forma, a psicometria ou medida em Psicologia se insere dentro da teoria da medida em geral que, por sua vez, desenvolve uma discussão epistemológica em torno da utilização do símbolo matemático (o número) no estudo científico dos fenômenos naturais. Trata-se, portanto, de uma sobreposição, ou melhor, de uma interface entre sistemas teóricos de saber diferentes, tendo a teoria da medida a função de justificar e explicar o sentido que tal interface possui.

CIÊNCIA E MATEMÁTICA

A Matemática e a Ciência empírica são sistemas teóricos (ou de conhecimento) muito distintos e, em termos estruturais, não são comensuráveis. Na verdade, os dois sistemas têm objetos e metodologias próprios, distintos e irreversíveis entre si. Pode-se discernir esta distinção atentando para a Tabela 1-1. Observa-se que em nenhum momento ou sob nenhum critério os dois sistemas se assemelham estruturalmente. A Ciência tem como referente ou objeto os fenômenos da realidade, ao passo que a Matemática estuda como seu objeto o símbolo numérico (que é um conceito e não uma realidade empírica e nem uma propriedade desta realidade - Frege, 1884); a metodologia da Ciência é a observação sistemática e a da Matemática é a dedução; o critério de verdade para a Ciência é o teste empírico, ao passo que para a Matemática é a consistência interna do argumento.

Assim, a primeira afirmação, no contexto da teoria da medida, consiste em dizer que o sistema científico do conhecimento não tem nada a

ver com a Matemática e vice-versa, falando-se em termos das estruturas epistemológicas dos dois saberes. O mesmo tipo de argumentação pode ser feito da Ciencia com relação aos outros sistemas de saber (Filosofia, Teologia, etc.).

Tabela 1-1 — Enfoque epistemológico de Ciência e Matemática

SISTEMA TEÓRICO	OBJETO	ATITUDE	METODOLOGIA	VERDADE	CERTEZA	CRITÉRIO DE VERDADE
Ciência (empírica)	Fenômenos naturais	Empírica	Observação e Controle	Fato	Relativa	Teste Empírico
Matemática	Símbolo numérico	Transcendental	Dedução	Teorema	Absoluta	Consistência interna do argumento

A NATUREZA DA MEDIDA

Apesar dessa distância epistemológica entre Ciência e Matemática, a Ciência empírica se apercebeu das vantagens consideráveis que ela pode obter ao se utilizar da linguagem da Matemática para descrever o seu objeto próprio de estudo. Na verdade, se o modelo matemático não dita e nem fundamenta o conhecimento científico, parece que é o uso deste modelo que vem possibilitando distinguir níveis de progresso no conhecimento científico. Esta afirmação, pelo menos, aparece claramente demonstrada na ciência da Física que, com o uso do modelo matemático, pôde passar de um estágio pré-histórico à Física moderna. Além disso, "Os instrumentos e técnicas de medida propiciam a ponte mais útil entre os mundos do dia-a-dia do leigo e dos especialistas em ciência" (Klein, 1974, p.24).

O uso do número na descrição dos fenômenos naturais constitui o objeto da teoria da medida. Esta teoria está razoavelmente axiomatizada somente nas ciências físicas, aparecendo ainda lacunar nas ciências sociais e do comportamento, onde, aliás, ainda se discute a viabilidade epistemológica da própria medida.

A natureza da medida implica alguns problemas básicos, dentre os quais três devem ser mencionados (Luce, Suppes, 1986; Suppes, Zinnes, 1963; Campbell, 1928, 1938): a representação, a unicidade e o erro.

O problema da representação ou o isomorfismo

O problema central da medida consiste em justificar a legitimidade de se passar de procedimentos e operações empíricos (a observação) para uma representação numérica destes procedimentos. É justificável

designar ou expressar objetos ou fenômenos naturais através de números? Sim, se nesta designação se salvarem tanto as propriedades estruturais do número quanto as características próprias dos atributos dos fenômenos empíricos. Trata-se do teorema da representação.

O problema da unicidade da representação

Este problema diz respeito ao fato de que uma representação 'x' de um atributo natural seria a única viável e possível ou que a tal representação é a mais adequada que se possa imaginar para descrever o atributo em questão. Assim, por exemplo, a massa das coisas não pode ser descrita melhor ou pode ser descrita somente em termos de peso, este expresso em termos de quilogramas: não haveria outra maneira possível de se descrever massa. O teorema da unicidade define o nível de escala de medida; em outras palavras, ele define se a escala obtida será ordinal, intervalar, etc.

O problema do erro

A observação dos fenômenos empíricos é sempre sujeita a erros devidos tanto ao instrumental de observação (os sentidos e suas extensões através de instrumentos tecnológicos) quanto a diferenças individuais do observador, além de erros aleatórios, sem causas identificáveis. Assim, tipicamente toda e qualquer medida vem acompanhada de erros e, por conseqüência, o número que descreve um fenômeno empírico deve vir acompanhado de algum indicador do erro provável, o qual será analisado dentro de teorias estatísticas para determinar se o valor encontrado e que descreve o atributo empírico está dentro dos limites de aceitabilidade de medida.

A BASE AXIOMÁTICA DA MEDIDA

Há legitimidade no uso do número na descrição dos fenômenos naturais se e somente se as propriedades estruturais do mesmo forem salvaguardadas neste procedimento. Isto é, deverá haver isomorfismo estrito (relação de 1 para 1) entre propriedades do número e aspectos dos atributos da realidade empírica.

São propriedades básicas do sistema numérico a identidade, a ordem e a aditividade. A medida deve salvar, pelo menos, as duas primeiras destas propriedades — de preferência, as três.

Para melhor enquadrar a psicometria ou a medida em geral em ciências do comportamento neste contexto, a base axiomática da medida será melhor feita dentro das ciências físicas, fazendo em seguida as ressalvas e correções necessárias para o caso da medida em ciências sociais e do comportamento e, em especial, da Psicologia.

Axiomas do sistema numérico

Stevens (1951) sumariza o sistema numérico na Figura 1-1:
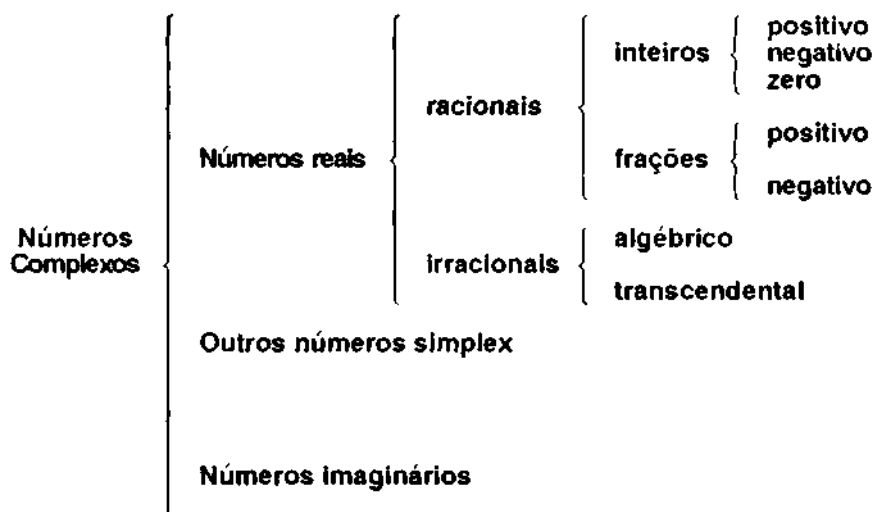
Fig. 1-1 — O sistema numérico

Estes vários tipos de números surgiram em épocas históricas diferentes, segundo as necessidades dos estudiosos e as necessidades da vida prática. Inicialmente só havia os números inteiros, que eram suficientes para contar objetos discretos, razão pela qual eles também são chamados de números naturais. Com eles se podia fazer as operações de adição e de multiplicação. Eles não davam sempre certo, porém, quando da subtração, especialmente quando se queria subtrair um número maior de um número menor. Esta limitação do sistema de inteiros fez com que o sistema fosse estendido para incluir números negativos e o zero. Com a divisão, o sistema de inteiros se mostrava ainda mais limitado, o que forçou a adoção de números fracionários. Este conjunto de números (inteiros positivos, negativos, zero e frações) constitui o sistema de números racionais, dado que qualquer número deste sistema pode ser expresso em termos de razão entre dois números inteiros. Excetuada a divisão por zero, todas as operações são possíveis dentro deste sistema numérico. Contudo, certas operações matemáticas não eram viáveis dentro do sistema, como, por exemplo, a raiz quadrada de 2. Inventaram-se, então, os números irracionais, e assim se fechou o círculo dos números reais, suficientes para permitir qualquer sorte de medida da realidade, até o presente.

A Matemática é um saber baseado em puras convenções; assim, tanto o seu objeto (o número) quanto suas regras sao convencionadas. As regras do jogo dos números se caracterizam por três grandes propriedades: identidade, ordem e aditividade.

Identidade (=)

Esta propriedade define o conceito de igualdade, isto é, que um número é idêntico a si mesmo e somente a si mesmo. Ela apresenta três axiomas (postulados aceitos e nao provados) que expressam a relação de IGUALAW:

— reflexividade: a = a ou a ≠ b. Números são idênticos ou são diferentes;
— simetria: se a = b, então b = a;
— transitividade: se a = b e b = c, então a = c. Duas coisas iguais a uma terceira são iguais entre si.

Ordem (>)

Esta propriedade se baseia na desigualdade dos números. Todo número é diferente de outro. Assim, excetuado o caso de igualdade, os números podem ser colocados numa seqüência invariável ao longo de uma escala linear: seqüência monotônica crescente. Também tem três axiomas, que expressam NAO IGUAL A (>):

— assimetria: se a > b, então b > a. A ordem dos termos não pode ser invertida;
— transitividade: se a > b e b > c, então a > c;
— conectividade: ou a > b ou b > a;

Um quarto axioma seria o de ordem-denso: números racionais são tais que entre dois números inteiros quaisquer há sempre um número racional; o intervalo entre dois inteiros não é vazio.

Aditividade (+)

Os números podem ser somados. Isto é, as quatro operações — adição, subtração, multiplicação e divisão (as três últimas sao redutíveis à primeira) — podem ser aplicadas aos números. Dois axiomas:

— comutatividade: a + b = b + a. A ordem dos termos não altera o resultado da adição;
— associatividade: (a + b) + c = a + (b + c). A ordem de associação ou de combinação dos termos não afeta o resultado.

Axiomas da medida

Como a medida consiste na atribuição de números às propriedades das coisas segundo certas regras, ela deve garantir que as operações empíricas salvem os axiomas dos números. A medida que salva todos esses axiomas é a mais sofisticada possível e, por isso, rara (escala de razão). A maioria das medidas, ao menos em ciências do comportamento,

se dão por satisfeitas se puderem salvar, pelo menos, os axiomas de ordem. Se somente os axiomas de identidade forem salvos (escala nominal), a operação propriamente não chega a ser medida, mas trata-se apenas de classificação, pois a única característica do número salva é a sua identidade; isto é, o número utilizado para uma operação empírica deve ser diferente do de uma outra operação. Para tanto, aliás, o número é utilizado tão-sòmente como numeral, a saber, um rabisco diferente de outro, que poderia ser substituído por qualquer outro sinal ou rabisco (desde que diferentes entre si) sem a menor conseqüência para a medida. O número, neste caso, serve apenas de etiqueta de uma classe de coisas. A medida realmente acontece quando se salvam, pelo menos, os axiomas de ordem dos números. Então, fica a pergunta: É possível se demonstrar a existência de ordem de magnitude nos atributos das coisas? Isto é, as coisas têm dimensões, entendidas estas como atributos mensuráveis, propriedades empíricas possuidoras de magnitude? Como resposta a esta questão poder-se-ia simplesmente assumir que sim: os atributos empíricos têm magnitude, como o senso comum nos parece dizer quotidianamente quando fala de 'mais do que', 'maior que' e expressões similares. Contudo, esta não parece ser uma base muito segura para fundamentar uma teoria da medida. É preciso, então, demonstrar empiricamente que tal ocorrência existe na realidade das coisas. Nas ciências físicas esta questão parece resolvida, mas nas ciências sociais e do comportamento ela ainda suscita acirradas controvérsias. Segue uma tentativa de demonstração experimental de axiomas da medida.

Demonstração empírica dos axiomas de ordem (Guilford, 1954)

Estes afirmam que, na medida, a ordem dada pelos números atribuídos aos objetos (transitividade e conectividade) deve ser a mesma obtida pela ordenação empírica destes mesmos objetos. Existe ordem ('maior que') nas propriedades das coisas. Exemplos: 1) de um metal que arranha um outro e não pode ser arranhado por este, diz-se que é mais duro — assim, uma ordem empírica de dureza pode ser estabelecida a partir da operação empírica de arranhar; 2) igualmente, o alinhamento de linhas mostra que uma é maior que outra, donde uma ordenação de objetos em termos de comprimento poder ser montada; 3) se um tom é dito mais alto que outro por uma amostra de sujeitos, diz-se que ele é mais agudo — assim, uma ordem de altura de tons *(pitch)* pode ser estabelecida; 4) se um sujeito resolve corretamente maior número de uma série de problemas do que outro, diz-se que ele é mais inteligente — assim, pode-se estabelecer uma escala de inteligência. As inversões que ocorrem são consideradas 'erros de medida' ou de observação, que devem ser tratados dentro da teoria da consistência, a qual visa mostrar que, apesar desses erros, há consistência na medida.

Demonstração empírica dos axiomas de aditividade

Esta demonstração parece ser possível somente no caso dos atributos extensivos, como massa, comprimento e duração temporal, bem como no caso da probabilidade. A aditividade se baseia na idéia de concatenação: a combinação (concatenação) de dois objetos ou eventos produz um terceiro objeto ou evento com as mesmas propriedades dos dois, mas em grau maior. Assim, tomando-se um objeto de comprimento 'x' (medido por uma unidade de comprimento qualquer, o metro, por exemplo), encontra-se um outro objeto com o mesmo comprimento 'x' que, junto (concatenado) ao primeiro, resulta num objeto maior 'z' com comprimento duas vezes o comprimento dos objetos individuais. O conceito de concatenação implica que A con B (A concatenando B) = A + B.

FORMAS E UNIDADES DE MEDIDA

Se cada atributo da realidade empírica apresentasse uma unidade-base natural específica de magnitude, a medida dele seria uma tarefa relativamente fácil. Seria suficiente verificar quantas unidades-base ele possui, e o número de unidades seria a medida do atributo em questão. Acontece, porém, que nem no mundo da Física todos os atributos permitem uma definição de unidade-base natural específica, como, por exemplo, no caso da velocidade. Disto resulta que deve haver mais de uma forma de se proceder à medida dos atributos da realidade que não seja a simples enumeração do número de unidades que o objeto apresenta.

Formas de medida

Hã diferentes maneiras (formas) de se atribuir números às propriedades dos objetos. Uma das taxonomías mais úteis consiste em distinguir três formas diferentes de mensuração: medida fundamental, medida derivada e medida por teoria (esta chamada de medida *by fiat* por Campbell, 1928, 1938). Pode-se igualmente falar em medida direta e medida indireta; e há outras ainda. A primeira, contudo, parece mais esclarecedora.

Medida fundamental

É a medida de atributos para os quais, além de se poder estabelecer uma unidade-base natural específica, existe uma representação extensiva. São dimensões (atributos mensuráveis) que permitem a concatenação, isto é, dois objetos podem ser associados, concatenados, formando um terceiro objeto de mesma natureza. Tal situação ocorre com os atributos de massa, comprimento e duração temporal. Estes atributos permitem uma medida direta e fundamental, dado que o instrumento utilizado para medi-los possui a mesma qualidade que se quer medir neles. Assim, ao se medir o comprimento de um objeto, utiliza-se um instrumento composto de

unidades de comprimento. A medida dele será dada pela coincidência de pontos entre o comprimento do objeto e a unidade de comprimento marcada no instrumento, por exemplo o metro.

Mesmo podendo ser possível conceitualmente se proceder a uma medida fundamental nos casos mencionados, nem sempre isto é empiricamente factível. Por exemplo, como se faria uma medida fundamental de distâncias astronômicas ou subatômicas? Ou como se poderia medir fundamentalmente a massa de uma galáxia? Nestes casos e semelhantes é preciso recorrer a outras estratégias de medida, como a medida derivada.

Medida derivada

Muitos atributos da realidade nao permitem urna medida extensiva e, portanto, nenhuma medida fundamental é deles possível. Eles podem, contudo, ser medidos indiretamente através do estabelecimento de uma relação com medidas extensivas. Este procedimento depende da prova empírica de que estes atributos são afetados independentemente por dois ou mais componentes. Se estes componentes permitem medida fundamental, então pode-se obter uma medida derivada para aqueles atributos não-extensivos através de uma função de potência entre os componentes da qual o atributo em questão é afetado. De qualquer forma, uma tal medida é derivada se finalmente ela pode ser expressa em termos de medidas fundamentais. Por exemplo, sabe-se que a massa varia em função de volume e de densidade: massa = volume x densidade. Como a massa permite medida fundamental (peso, expresso em quilos) e o volume também (o cubo do comprimento = m^3), então a densidade, que não possui medida fundamental, pode ser medida indiretamente em função de massa e volume (quilos dividido por metros cúbicos = kg/m^3).

Deve-se notar que o fundamento da função existente entre os componentes constitui uma lei, isto é, deve ser um dado empiricamente demonstrado e não somente baseado em alguma teoria. Assim, a massa sendo determinada pelo volume e pela densidade é uma descoberta científica, uma lei, não uma hipótese. Entende-se, portanto, por medida derivada de um atributo aquela cujos componentes do atributo, estabelecidos por uma lei empírica, tenham finalmente dimensões extensivas.

Esta é uma parte da história. Fala-se também de medida derivada quando, embora os componentes da função não forem redutíveis, em última análise, a medidas fundamentais, eles apresentam, contudo, uma unidade-base natural e específica. Assim, pode-se medir de modo derivado um atributo não-extensivo se ele for passível de ser expresso em termos de outros atributos que possuem uma unidade-base, ainda que estes não tenham propriedade extensiva. Exemplo: luminância pode ser medida através da unidade-base de intensidade luminosa (a candela) e

pela distância (medida em metros); de fato, ela é lum = cd/m^2. O mesmo vale para resistência elétrica, força do campo elétrico e do campo magnético.

Medida por teoria

Há outros atributos da realidade — e é o caso de quase todos em ciências sociais e do comportamento — que, além de não se apresentarem como dimensões extensivas, portanto incapazes de medida fundamental, também não podem ser expressos em termos de componentes extensivos e não possuem uma unidade-base natural, não permitindo, portanto, medida derivada. Estes são mensurareis somente com base em leis e teorias científicas.

1) Medida por lei: quando uma lei for estabelecida empiricamente entre duas ou mais variáveis, a(s) constante(s) típica(s) do sistema pode(m) ser medida(s) indiretamente através da relação estabelecida entre estas variáveis, como é o caso da viscosidade em Física e a lei do reforço em Psicologia.

2) Medida por teoria: quando nem leis existem relacionando variáveis, pode-se recorrer a teorias que hipotetizam relações entre os atributos da realidade, permitindo assim a medida indireta de um atributo através de fenômenos a ele relacionados via teoria. O importante neste caso é garantir que haja instrumentos calibrados para medir (fundamental mente ou de outra forma válida) os fenômenos com os quais o atributo em questão esteja relacionado pela teoria. Mesmo em Física isto ocorre, como é o caso da medição das distâncias galácticas. Assim, medindo o movimento das linhas espectrais para o vermelho estar-se-ia medindo as distâncias astronômicas, dada a teoria de que existe uma relação siste mática entre a distância de uma galáxia e a velocidade do seu afastamen to. O mesmo vale para o efeito Doppler, que afirma (teoria) que uma luz que se afasta tende a espalhar as ondas, reduzindo sua freqüência.

Unidades de medida

Normalmente existe interdependência entre os fenômenos, de sorte que, ao se variar um deles, o outro covaria com ele. Esta covariancia pode ser expressa por alguma constante. Estas constantes podem ser universais, como o caso da gravitação universal que covaria com as gravitações locais de um sistema menor, por exemplo, a da massa, chamada inclusive de constante universal de gravitação. Outras constantes pertencem a algum sistema específico (constantes do sistema ou locais), como a constante entre massa e volume em Física ou as constantes da lei do reforço em Psicologia. Evidentemente, a descrição de tais constantes pode constituir uma medida indireta.

Além de constantes que relacionam dois ou mais atributos, os próprios atributos variam por conta própria, assumindo diferentes

magnitudes, isto é, eles são dimensões, entendendo por isso que eles podem variar de magnitude e, portanto, podem ser mensuráveis. Neste caso, seria extremamente útil se houvesse, para cada atributo diferente, uma unidade básica com a qual se pudesse determinar a magnitude do mesmo. De fato, qualquer unidade que se queira definir serve aos propósitos da medida, bastando haver consenso sobre a mesma. Mas é fácil ver as vantagens de se estabelecerem unidades-base aceitáveis para todos. Nas ciências físicas, este esforço tem sido constante. O critério que tem guiado os físicos na procura destas unidades-base foi a busca de um fenômeno natural de estabilidade máxima que pudesse servir como padrão físico da unidade-base para o sistema. A história da procura destas unidades tem lances de Babel, pois cada região do mundo tinha seus sistemas de medida, incomensuráveis com os de outras regiões. Há cerca de 200 anos, contudo, uma procura mais sistemática e mais entrosada em nível mundial tem sido desenvolvida até que culminasse no Système International des Unités (abreviado SI), definido na 11th General Conference on Weights and Measures (Paris, 1960), onde foram estabelecidas seis unidades-base ou primárias para os fenômenos físicos, sendo todas as restantes medidas derivadas destas seis primárias (Klein, 1974; Luce, Suppes, 1986). A Tabela 1-2, na página seguinte, sintetiza estas unidades-base consensuais, inclusive o mol, unidade posteriormente acrescida.

A grande maioria das outras unidades em Física são unidades derivadas destas seis unidades-base. Por exemplo, densidade é igual a peso por volume (kg/m^3), velocidade a metros por segundos (m/s), luminância a intensidade da luz por área que é expressa em termos de distância (cd/m^2), volt é watts por ampère (V=W/A), watt é joule por segundo (W=J/s), joule é newton vezes comprimento (J = N.m), newton é peso vezes distância por tempo ($N=kg'm/s^2$), etc.

A procura de unidades similares em ciências sociais e do comportamento é algo ainda precário, exceto onde medidas fundamentais forem possíveis, como talvez em psicofisica (medida dos estímulos) e na análise experimental do comportamento (medidas de estímulos e freqüência de respostas).

A MEDIDA EM CIÊNCIAS SOCIAIS E DO COMPORTAMENTO

Medidas fundamentais nestas ciências parece difícil de serem concebidas. Mesmo em Economia, que se apresenta como a mais desenvolvida nesta área, parece ter caído em descrédito a concepção de que a escolha dos sujeitos se reduziria à avaliação da quantidade e preço dos bens. De fato, há ali fatores subjetivos que co-determinam a escolha dos sujeitos, fatores agrupados sob o construto de utilidade. Também não parece aceitável que a utilidade de um conjunto de bens possa ser reduzida à soma das utilidades individuais destes bens. Deste problema

Tabela 1-2 — Unidades-base da Física

ATRIBUTO	UNIDADE	SIGLA	PADRÃO FÍSICO (Definição do SI)
Comprimento	metro	m	"O metro e o comprimento igual a 1.650.763.63 comprimentos de onda no vácuo da radiação correspondente à transição entre os níveis 2 p10 e 5 d5 do átomo do Criptônio-86."
Massa	quilograma	kg	"O quilograma (unidade de massa) é a massa de um cilindro especial feito de liga de platina e de iridio, que é considerado como o protòtipo internacional do quilograma. e é conservado sob os cuidados do Bureau International des Poids et Mesures num cofre forte em Sèvres, França."
Tempo	segundo	s	"O segundo é a duração de 9.192.631.770 períodos (ou ciclos) da radiação correspondente à transição entre dois níveis hiperfinos do átomo de Césio-133."
Corrente elétrica	ampère	A	"O ampère, unidade de corrente elétrica, é a corrente constante que, se mantida em dois condutores paralelos de comprimento infinito. de uma grossura negligível, e colocados a 1 metro de distância num vácuo, produzirá, entre estes condutores, uma força igual a 2 X 10 ^{7}N por metro de comprimento (cerca de 0,1kg)."
Temperatura	kelvin	K	"O kelvin, a unidade de temperatura ter-modinâmica. ó atração 1/273,16 da temperatura termodinâmica do triplo ponto da água (no qual gelo. água e vapor estão em equilíbrio — igual a-273,16°C)."
Intensidade da luz	candela	cd	"Luminosidade de 1/600.000 de um metro qua-drado de pura platina fundida no ponto de se solidificar. Isto corresponde a uma temperatura de 2.045ºK."
Massa atômica	mole	mol	Montante de substância que corresponde á somados pesos atômicos de todos os átomos que compõem uma molécula.

surgiu a teoria moderna da utilidade baseada na teoria dos jogos. Em psicofisica também se tenta enquadrar a medida como sendo fundamental. Entretanto, para tanto dever-se-ia modificar a definição de medida fundamental como sendo a medida de atributos extensivos. Em psicofisica

o atributo de interesse é a resposta do sujeito a estímulos físicos. Estes certamente podem permitir medida fundamental, mas não são eles o interesse específico direto da medida psicofisica, mas sim a resposta a eles. E desta nao há como visualizar uma medida fundamental, dado que não é um atributo extensivo. A medida da resposta se faz em função da sua relação com o estímulo, relação estabelecida por uma lei empiricamente demonstrada. A medida, portanto, se baseia numa função entre "componentes". Componentes está entre aspas porque o estímulo realmente não é componente da resposta no sentido dado nas medidas derivadas, nas quais os componentes relacionados são propriedades constituintes do atributo medido derivadamente, como massa em função de volume e densidade.

Se medida fundamental não é defensável em ciências sociais e do comportamento, nem a derivada o é. Resta, então, a possibilidade de se medir nestas ciências por uma terceira forma, que vimos apresentando sob a égide de medida por teoria, que congrega aquelas formas de medida não redutíveis finalmente a medidas fundamentais. Duas formas de medida são aqui destacáveis: medida por lei e medida por teoria propriamente. As duas podem ser enquadradas sob medida por teoria, dado que a lei constitui uma hipótese derivável de alguma teoria e empiricamente demonstrável.

Medida por lei

A medida por lei é comum nas ciências sociais e do comportamento. Em Psicologia, em particular, ela é a história na psicofisica e na análise experimental do comportamento. Em psicofisica, a história que vai de Weber a Stevens é a medida por lei: lei da constante (Weber), lei logaritmica (Fechner) e lei da potência (Stevens). Na análise experimental do comportamento temos as várias leis do reforço, por exemplo.

Em que consiste uma medida por lei? Mede-se por lei quando se quer demonstrar empiricamente que dois ou mais atributos estruturalmente diferentes mantêm entre si relações sistemáticas. Duas condições são expressas nesta concepção: 1) os atributos sao de natureza diferente, um não é redutível ao outro — por exemplo, a cor e a distância são dois atributos distintos dos fenômenos físicos no caso do desvio para o vermelho das linhas espectrais dos objetos na medida de distâncias, do mesmo modo que a resposta e o estímulo são dois atributos diferentes no caso da medida psicofisica e da análise experimental do comportamento; 2) uma relação sistemática foi demonstrada cientificamente (empiricamente) existir entre estes atributos. Assim, as manipulações efetuadas num atributo repercutem sistematicamente no outro, donde é possível estabelecer uma função de covariancia entre os dois, uma lei.

Medida por teoria

Uma teoria não é uma lei, dado que ela é composta de axiomas ou postulados e nao de fatos empíricos. Ademais, ela é científica se de seus axiomas é possível deduzir hipóteses empiricamente testáveis. O caso da medida por teoria ocorre também em Física, como ficou dito acima. No caso das ciências sociais e do comportamento, ela é mais facilmente discernível quando se trabalha com variáveis hipotéticas, isto é, nas teorias da estrutura latente ou da modelagem latente *(latent modeling).*

Nesta concepção, a teoria versa sobre processos mentais (estruturas psicológicas hipotéticas), conceitualizando sua estrutura e sua dinâmica, e define o conjunto de comportamentos que os expressa. Assim, uma teoria da inteligência trata dos processos cognitivos, sua estrutura, sua gênese, seu desenvolvimento, seus processos operativos, etc, bem como dos comportamentos típicos em que ela se expressa, tais como resolver problemas numéricos, problemas espaciais, problemas com palavras, etc. E é no nível dos comportamentos que se faz a medida. De sorte que, ao se medir por teoria neste caso, consiste em dizer que ao se proceder a medida de um atributo empírico (o comportamento) está-se de fato medindo a estrutura psicológica latente, isto é, um outro atributo e de outra natureza (um atributo hipotético). A função da teoria é fundamentalmente garantir a legitimidade de tal operação; trata-se, portanto, de um problema de representação: a operação empírica de medir o comportamento representa a medida da estrutura latente. Legitimar tal representação é a função da teoria psicométrica (a questão da validade psicométrica da medida).

Na verdade, pode-se ver três grandes linhas da teoria da medida por teoria: a teoria da detecção do sinal em psicofisica, a teoria dos jogos em Psicologia Social e a teoria das escalas e testes (a psicometria, num sentido mais restrito). O fato de trabalharem com construios hipotéticos (processos cognitivos, preferências, utilidade, etc.) seria o elo entre estes tipos de teorias, bem como a procura de formas de medida desses mesmos construios.

A teoria da detecção do sinal trabalha com dois parâmetros: relação sinal-ruído ('d') e a disposição do sujeito ('beta'). O primeiro parâmetro define o grau de detectabilidade do sinal contra um fundo de ruído, e o 'beta' define o nível de vontade ou disposição que o sujeito tem de ver o sinal quando ele está presente.

A teoria dos jogos trabalha com o conceito de utilidade, introduzido pela ciência econômica. Esta variável representa a preferência do sujeito na escolha de uma alternativa de ação dentre duas ou várias. A teoria dos jogos determina as probabilidades de ganhos e perdas associadas com a

escolha de cada alternativa, bem corno permite verificar o nível de preferência do sujeito com relação a cada alternativa, possibilitando, assim, uma ordenação destas alternativas em termos de preferência.

A psicometria, em sentido estrito, trata da medida de construtos psicológicos ou de traços latentes, através de comportamentos verbais ou motores que seriam a representação daqueles traços. Os parâmetros com que lida a psicometria são a legitimidade de tal representação (validade) e a análise dos itens comportamentais em termos de dificuldade e discriminação (Item Characteristic Curve — ICC, Lord, 1980).

NÍVEIS DA MEDIDA (ESCALAS DE MEDIDA)

Dependendo da quantidade de axiomas do número que a medida salva, resultam vários níveis de medida, as chamadas escalas de medida. Como vimos, são três os axiomas básicos do número: identidade, ordem e aditividade. O último apresenta dois aspectos úteis para o presente problema: origem e intervalo ou distância. Quanto mais axiomas do número a medida salvaguardar, maior será o seu nível, isto é, mais ela se aproxima da escala numérica ou métrica e maior será o isomorfismo entre o número e as operações empíricas. Assim, podemos considerar cinco elementos numéricos para definir o nível da medida: identidade, ordem, intervalo, origem e unidade de medida. Destes cinco elementos, os mais discriminativos dos níveis são a origem e o intervalo, dado que a ordem é uma condição necessária para que realmente haja medida. Se a medida somente salva a identidade do número, na verdade não se trata de medida, mas sim de classificação e contagem. Neste caso (escala nominal), os números não são atribuídos a atributos dos objetos, mas o próprio objeto é identificado por rótulo numérico. Este rótulo nem precisaria ser numérico, dado que não importa que símbolo ou rabisco pode ser utilizado com a mesma função de distinguir objetos um do outro ou classe de objetos de outra classe. A única condição necessária é que se salvaguarde a identidade do símbolo, isto é, um mesmo símbolo não pode ser duplicado para identificar objetos diferentes, como também diferentes símbolos não podem ser usados para identificar objetos idênticos. Embora não estejamos neste caso medindo, a escala numérica que resulta desta rotulação adquire direito ao nome escala, dado que ela corresponde em parte à definição de medida que reza "medir é atribuir números às coisas empíricas".

O esquema a seguir ilustra como se originam as várias escalas de medida, enquanto que a Tabela 1-3, por outro lado, sintetiza as características de cada escala.

ORIGEM

INTERVALO		NAO-NATURAL	NATURAL
	NAO-IGUAL	ORDINAL	ORDINAL
	IGUAL	INTERVALAR	RAZÃO

Uma escala numérica pode ser transformada numa outra equivalente, se forem respeitados os elementos da invariância nesta transformação. Uma escala de maior nível pode utilizar as operações estatísticas de uma escala inferior, mas perde informação dado que as estatísticas próprias de uma escala inferior são menos eficientes, isto é, são menos robustas. Não é permitido (é erro) utilizar estatísticas de uma escala de nível superior numa inferior, dado que esta não satisfaz os requisitos necessários para se utilizar de procedimentos estatísticos superiores. São chamados de paramétricos os procedimentos estatísticos da escala intervalar, porque nela os números possuem caráter métrico, isto é, são adicionáveis, enquanto os não-paramétricos não são métricos, visto que representam somente postos e não quantidades somáveis.

Tabela 1-3 — Características das escalas numéricas de medida.

ESCALA	AXIOMAS SALVOS	INVARIÂNCIAS	LIBERDADES	TRANSFORMAÇÕES PERMITIDAS	ESTATÍSTICAS APROPRIADAS
Nominal	identidade		ordem intervalo origem unidade	Permutação (troca 1 por 1)	Freqüências: f,%,p,Mo,X^2,C
Ordinal	identidade ordem	ordem	intervalo origem unidade	Monotônica crescente (isotonia)	Não-paramétricas: Md, r_s U, etc.
Intervalar	identidade ordem actividade	ordem intervalo	origem unidade	Linear de tipo y = a + bx	Paramétricas: M,DP,r,t,f,etc.
Razão	identidade ordem aditividade	ordem intervalo origem	unidade	Linear de tipo y= bx (similaridade)	Média geométrica, Coef. variação, Logaritmos

Conceito de erro

A medida é um procedimento empírico, e não existe procedimento empírico isento de erro. Esta não é uma afirmação lógica, mas pode ser considerada um postulado e empiricamente verificada através de operações de mensuração. Mesmo na medida fundamental, é impossível evitar-se o erro. Argumentando com Popper (1972), podemos dizer que medir consiste na determinação da coincidência de pontos: um sinal no objeto a ser medido e um sinal no instrumento de medida (metro, por exemplo). Agora, não existe tal coincidência no sentido de que os dois pontos se fundem num ponto único; há apenas uma justaposição dos dois pontos. A precisão perfeita da justaposição só seria finalmente efetuada se pudesse ser verificada num aumento ao infinito desses dois pontos, pois que, com o aumento deles, se verifica que os pontos realmente (de fato) não estão perfeitamente alinhados, mas apenas aparecem mais ou menos próximos. Assim, a coincidência se faz dentro de um intervalo: o ponto do corpo medido cai dentro de um intervalo de pontos no instrumento (extremos de condensação). Quanto menor este intervalo, maior a precisão da medida. Por esta razão, é costumeiro entre os cientistas apresentar, além do valor da medida, o seu equivalente erro provável, o qual define precisamente estes extremos de condensação.

Tipos de erro

Os erros podem ser debitados ou à própria observação ou à amostragem de objetos ou eventos na qual a medida foi realizada.

Erros de observação

Há quatro fontes principais de erros de observação: erros instrumentais, devidos a inadequações do instrumento de observação; erros pessoais, devidos às diferentes maneiras de cada pessoa reagir; erros sistemáticos, devidos a algum fator sistemático não controlado, como, por exemplo, medir a temperatura a um nível diferente do nível do mar; e erros aleatórios, sem causa conhecida ou conhecível. Há, inclusive, curiosos acontecimentos neste particular, como a demissão do seu assistente pelo astrônomo real Nevil Maskelyne (Inglaterra) porter observado a passagem de estrelas e planetas meio segundo depois do que tinha ele mesmo observado. O problema não é tanto a existência desses erros, que são inevitáveis, mas sim identificar as suas fontes e propor meios de reduzi-los. A Tabela 1-4 dá uma síntese desta problemática.

Tabela 1-4 — Erros de medida: fontes e controle

TIPO	CAUSA	CONTROLE
instrumental	instrumento	calibração
pessoal (observador)	diferenças individuais	atenção, treinamento
sistemático	fator específico	experimental ou estatístico
aleatório	não conhecida	teorias do erro (probabilidade)
amostragem	coleta da amostra	representatividade da amostra (teoria estatística)

Erros de amostragem

Como a pesquisa empírica normalmente não pode ser feita sobre todos os membros de uma população de eventos ou objetos, tipicamente se seleciona uma amostra destes eventos ou objetos. Esta escolha de indivíduos no meio de uma população é sujeita a desvios, vieses, isto é, erros. O problema não são os erros em si, se o interesse fosse tirar conclusões sobre a amostra selecionada. Acontece, porém, que o interesse do pesquisador é tirar conclusões ou fazer inferências sobre toda a população da qual a amostra foi retirada. Neste caso, o erro de amostragem é desastroso, uma vez que poderia ocasionar inferências errôneas, dada a presença de vieses da amostra com respeito a esta população (falta de representatividade). Para solucionaros problemas advindos da seleção da amostra, foi desenvolvida a teoria estatística da amostragem.

A teoria do erro

Dado que o erro está sempre presente em qualquer medida e que sua presença constitui uma ameaça séria à tomada de decisões científicas, é de capital importância que haja meios de neutralizar ou diminuir os seus efeitos ou, pelo menos, de conhecer sua grandeza, o mais aproximado possível, para saber o tamanho do risco em que se está incorrendo ao tomar decisões baseadas na medida. Todos os esforços para controlar o erro através de procedimentos experimentais são necessários, mas nem por isso o erro vai desaparecer, visto que a sua ocorrência é imprevisível, isto é, não é nunca possível se determinar as causas de todos os erros possíveis numa medida. Para enfrentar esta situação foi desenvolvida a teoria do erro, baseada na teoria da probabilidade e dos eventos casualóides.

Um evento casualóide ou aleatório é definido por Popper (1974, p. 190): "Uma seqüência-evento ou seqüência-propriedade, especialmente uma alternativa, se diz 'casualóide' ou 'aleatória' se e somente se os limites das freqüências de suas propriedades primárias forem 'absoluta-

mente livres', isto é, indiferentes a qualquer seleção que se apoie nas propriedades de qualquer ênupla de predecessores". Em palavras mais simples, um evento empírico é aleatório se sua ocorrência não pode ser predita a partir dos eventos que ocorreram antes dele, isto é, ele é totalmente independente (livre) com relação ao que aconteceu antes. Imagine o jogo de lançar uma moeda para obter cara ou coroa ou o de um dado: qualquer que tenha sido o resultado nos lançamentos anteriores do dado, o resultado (um entre os seis possíveis) do próximo lançamento é totalmente imprevisível — isto é liberdade absoluta.

O erro na medida é considerado um evento aleatório, pela teoria do erro. Feita esta suposição, então é possível tratar o erro dentro da teoria da probabilidade, do teorema de Bernoulli, que baseia a lei dos grandes números e da curva normal, que determina a probabilidade de ocorrência dos vários elementos da série, no nosso caso, da série aleatória composta dos vários tamanhos de erros cometidos na medida.

A curva normal define que uma seqüência aleatória de eventos empíricos se distribui normalmente em tomo de um ponto modal (média) igual a 0 e uma variância igual a 1. Este valor modal, no caso de uma distribuição de erros, significa que estes se cancelam no final, dado que este valor (0) é o que possui a maior probabilidade na distribuição. Contudo, isto é absolutamente verdadeiro somente na distribuição de uma série aleatória de um número infinito de eventos, segundo o teorema de Bernoulli. Este teorema, na verdade, afirma que um segmento 'x' de elementos de uma série aleatória infinita 'A' (isto é, com liberdade absoluta) que se aproxima da série total (x->A) possui os mesmos parâmetros desta série. Isto significa que, quanto maior o segmento, mais próximo está dos parâmetros da série ou, em outras palavras, quanto maior o segmento, menor o desvio dos parâmetros dele dos da série. Diz Popper (1974, p.198): "Assim, o teorema de Bernoulli assevera que os segmentos mais curtos de seqüências casualóides mostram, muitas vezes, grandes flutuações, enquanto que os segmentos longos sempre se comportam de modo que sugerem constância ou convergência; diz o teorema, em suma, que encontramos desordem e aleatoriedade no pequeno, ordem e constância no grande. É a este comportamento que se refere a expressão 'lei dos grandes números".

Na prática da pesquisa, contudo, o erro da medida é expresso pelo erro padrão da medida, que é o valor médio da variância, isto é,

$$EP = \frac{\sqrt{\sigma^2}}{\sqrt{N-1}} = \frac{DP}{\sqrt{N-1}}$$

A informação dada pelo erro padrão da medida esclarece que a medida verdadeira de um atributo se situa entre o valor médio das medidas

efetuadas e um erro padrão em torno dele (isto é, mais um erro padrão e menos um erro padrão).

IMPORTÂNCIA DA MEDIDA

Poder-se-ia perguntar, diante de tantas dificuldades que a medida apresenta, se há vantagem em se utilizar métodos de medições em lugar de métodos puramente qualitativos ou descritivos. Parece que a resposta deva ser positiva, porque aqueles métodos se apresentam superiores a estes em, pelo menos, duas áreas: precisão e simulação.

Precisão

Apesar de a medida nunca ser destituída de erro, ela é capaz de definir limites dentro dos quais os reais valores dos atributos medidos se encontram. O conceito de pontos de condensação ou de extremos imprecisos (Popper, 1974) nos indica a solução da questão da precisão da medida. Fazer pontos coincidirem (ponto extremo do atributo do objeto a ser medido e ponto de referência do instrumento de medida) significa determinar que o ponto do atributo cai dentro de um intervalo de pontos extremos do instrumento. A questão, então, se reduziria a determinar estes pontos extremos do intervalo, que, por sua vez, também caem dentro de um intervalo cujos pontos extremos precisariam ser determinados, e assim indefinidamente; isto é, nunca daria para decidir nenhum intervalo de pontos de condensação. Entretanto, os pontos extremos do intervalo de condensação seriam definidos por intervalos cada vez menores, de sorte que se pode finalmente definir um intervalo, o menor possível, com pontos extremos imprecisos, dentro do qual o valor real do atributo se encontra. Assim, fica definido um intervalo mínimo mais provável dentro de seus pontos extremos e, igualmente, a margem de erro tolerada ou provável, de sorte que não se contentaria em simplesmente afirmar que o atributo é mais ou menos de tal magnitude, mas que ele tem uma magnitude definida dentro de limites (intervalo) assim definidos. A redução ao mínimo do intervalo dos pontos de condensação, evidentemente, depende de avanços tecnológicos no instrumental de medição.

Sendo isso possível, ficam mais precisas tanto a descrição do fenômeno natural quanto a comunicação sobre o mesmo. Fica também mais exata a definição das operações e procedimentos utilizados na observação dos mesmos fenômenos. A medição não torna a observação possível, mas a torna mais unívoca, isto é, menos ambígua, mais precisa. Esta vantagem da medição se torna ainda mais crucial na observação do muito grande (macroscópico) e do muito pequeno (microscópico).

A simulação

A manipulação da realidade é geralmente complexa, difícil e custosa. Além disso, às vezes ela é impossível ou eticamente condenável. Por exemplo, não parece aceitável querer estudar os efeitos da bomba atômica sobre uma cidade explodindo uma. Mas, conhecendo com precisão as relações entre os componentes em jogo e suas magnitudes, pode-se utilizar modelos matemáticos para simular os efeitos que queremos estudar e que, de outro modo, seria impossível ou impraticável pesquisar.

CONCLUSÃO

A medida em ciências empíricas não pode ser considerada uma panacéia para decidir todos os problemas do conhecimento da realidade, inclusive porque não é ela que define o objeto e nem o método da Ciência. Mas, diante das vantagens apresentadas, seria quiçá até irracional não se aproveitar da medida como instrumental de trabalho no estudo da realidade. A história da Ciência parece demonstrar, inclusive, que o avanço do conhecimento científico está ligado ao maior ou menor uso da medida, sobretudo quando ela está baseada numa teoria axiomatizada, isto é, quando há a explicitação clara do maior número possível dos axiomas necessários. Infelizmente, na medida em ciências sociais e do comportamento, esta axiomatização está longe de ser uma realidade. Mesmo assim, **a** discussão sobre a viabilidade da medida nestas ciências parece uma disputa mais inócua que produtiva; **uma** discussão de como se proceder à medida parece mais substantiva, produtiva e útil para o desenvolvimento destas ciências.

REFERÊNCIAS BIBLIOGRÁFICAS

CAMPBELL, N.R. *An account of the principles of measurement and calculations.* London: Longmans Green, 1928.

--------------. Symposium: Measurement and its importance for philosophy.*Proceedings Aristot. Society Suppl.,* London, v.17, p.121-142, 1938.

FREGE, J.G. Os fundamentos da aritmética: uma investigação lógico-matemática sobre o conceito de número. In: OS PENSADORES: Peirce e Frege. Sao Paulo: Victor Civita, 1984. p. 195-276.

GUILFORD, J.P. *Psychometric methods.* New York: McGraw-Hill, 1954.

KLEIN, H.A. *The world of measurements,* masterpieces, mysteries and muddles of metrology. New York: Simon and Schuster, 1974.

LORD, F.M. *Applications of item response theory to practical testing problems.* Hillsdale (NJ): Erlbaum, 1980.

LUCE, R.D., SUPPES, P. Measurement theory.In: THE NEW ENCYCLOPAEDIA BRITANNICA. London: Britannica, 1986. v.23, p.792-798.

POPPER, K.R. *A lógica da pesquisa científica.* Sao Paulo: Cultrix, 1972.

--------------. *Conjeturas e refutações.* Brasília: Ed. UnB, 1974.

STEVENS, S.S. Mathematics, measurement and psychophysics. In: STEVENS, S.S. (Ed.). *Handbook of experimental psychology.* New York: Wiley, 1951. p.1-49.

SUPPES, P., ZINNES, J.L. Basic measurement theory. In: LUCE, R.D., BUSH, R.R.,GALANTER, E.G. (Eds.). *Handbook of mathematical psychology.* New York: Wiley, 1963. v.I, p.1-76.

CAPÍTULO 2

TEORIA E MÉTODOS EM PSICOFÍSICA*

Maria Angela Guimarães Feitosa
Instituto de Psicologia
Universidade de Brasilia

Informalmente, a psicofisica nasceu em aproximadamente 130 A.C., na Astronomia, nas mãos de Hiparco, com o desenvolvimento de uma escala para quantificação de magnitude estelar; formalmente, com a publicação do *Elemente der Psychophysik,* por Gustav Theodor Fechner, em 1860, como uma proposta metodológica para o estudo experimental da alma. Este evento foi precedido por alguns outros eventos relevantes que merecem menção. Daniel Bernoulli propôs, em 1738, que as pessoas não agem com base no valor real do dinheiro, mas em uma transformação psicológica do seu valor real. Bernoulli propôs que a utilidade do dinheiro aumenta como uma função logarítmica da quantidade do dinheiro. Em 1824, o matemático Herbart propôs o conceito de limiar, ao afirmar que eventos mentais precisam ter uma certa força para serem conscientemente experimentados. Herbart propôs ainda o uso de pressupostos e conceitos matemáticos para lidar com conceitos psicológicos. Weber (1834-1846), pesquisando sobre somatossentidos, observou que pesos grandes levantados precisam diferir mais do que pesos pequenos para serem percebidos como diferentes, levando à formulação de que a diferença apenas perceptível (DAP) é uma razão do peso levantado, criando as condições para que Fechner enunciasse a lei que mais tarde veio a ter o nome de Weber. Em torno de 1850, os astrônomos J. Herschel, W.R. Dawes e N.R. Pogson formularam, independentemente de Fechner, a lei psicofisica básica de Fechner, no contexto de desenvolver a escala de magnitude estelar originalmente concebida por Hiparco. Na década de 1850, o físico J.A.F. Plateau inventou o método de bissecção, no qual o observador, dados dois valores de estímulo, tinha que produzir um terceiro que parecesse ocupar a metade do intervalo entre os estímulos padrão. Especificamente, Plateau

* A preparação deste capitulo foi parcialmente financiada pelo CNPq, através do Auxílio Integrado 50.0493/91-0. A autora agradece a Luiz Pasquali, Miosotis Lúcio e Rosana Maria Tristão Taveira, pela leitura crítica de versão preliminar e pelas várias sugestões oferecidas. Miosotis Lúcio também colaborou no levantamento da bibliografia.

solicitou a artistas que, a partir de tintas branca e preta, produzissem um cinza que parecesse estar equidistante do branco e do preto.

A publicação do *Elemente der Psychophysik* teve impacto na comunidade científica européia e foi decisiva para o estabelecimento da própria Psicologia como disciplina científica. O estabelecimento e o progresso da psicofisica não foi, entretanto, ausente de percalços. A Universidade de Cambridge, na Inglaterra, recusou permissão para o estabelecimento de um laboratório de psicofisica porque isto "seria insultar a religião por colocar a alma humana em um par de escalas", uma reação que lembra a que sofreu Charles Darwin com o enunciado da teoria da evolução das espécies. Nos Estados Unidos, William James deixou publicamente claro seu desprezo pelo trabalho de Fechner, o que contribuiu para a pouca participação dos laboratórios americanos de Psicologia Experimental na consolidação da psicofisica (Adler, 1980; Boring, 1957).

A psicofisica criada por Fechner e os desdobramentos sob sua influência teórica e metodológica são referidos na literatura como psicofisica fechneriana ou clássica. Fechner tinha um interesse especial pela questão mente-corpo, e sua proposta psicofisica básica, R=f(S), tem embutidos o pressuposto de identidade psicofisica e o de que as pessoas não têm acesso direto às sensações, as quais deveriam ser medidas indiretamente através de sucessivos limiares diferenciais. Mais especificamente, ele pressupunha que a diferença apenas perceptível (DAP) é a unidade de sensação, que diferentes DAPs correspondem a uma mesma sensação e que as DAPs podem se somar para constituir uma escala. O investimento de Fechner no estudo das sensações não deve ser interpretado como refletindo sua posição sobre o objeto principal do desenvolvimento da psicofisica, mas como refletindo o fato de que Fechner viveu em uma época em que eram influentes filósofos associacionistas empiricistas que defendiam que todo o conhecimento e as idéias se devem, em última instância, a experiências sensoriais com o mundo externo. Não surpreende, portanto, que tendo Fechner um profundo interesse pela Filosofia, escolhesse as sensações como o ponto de partida de seu extenso "programa" psicofisico, o qual incluía o estudo da memória e do pensamento.

QUESTÕES CENTRAIS EM PSICOFÌSICA

Embora o objetivo principal deste livro seja metodologia de pesquisa em Psicologia, é impossível discorrer sobre métodos psicofísicos sem fazer uma incursão, ainda que breve, na teoria pertinente. Caso contrário, corre-se o risco de propiciar uso indevido do texto e de se passar ao leitor pouco mais do que um *cookbook,* porque métodos não costumam ser desenvolvidos dissociados de um referencial teórico ou de um problema cuja solução exija o desenvolvimento de uma estratégia

apropriada. Além disso, os métodos psicofísicos têm sido utilizados em campos tão distintos que a análise crítica da aplicabilidade do método específico não pode ser feita fora do contexto de pesquisa em que o método é usado. A organização das partes subseqüentes deste capítulo bem como os diversos exemplos escolhidos procuram destacar algumas destas questões. Associada à descrição de cada método, também será feita uma descrição básica do tratamento de dados proposto e consistente com os pressupostos teóricos subjacentes ao método. Tratamento sofisticado das teorias, dos métodos psicofísicos e da análise quantitativa dos dados pode ser encontrado em Falmagne (1985) e em Luce e Krumhansl (1987).

O objeto de estudo da psicofisica

Independentemente dos pressupostos e objetivos originais de Fechner, contemporaneamente a psicofisica se propõe a descrever as relações entre propriedades do mundo físico e a forma como as pessoas respondem a elas. O objeto central da psicofisica são as sensações, entendendo-se que sensação é um construto acerca dessas relações primitivas dos organismos com o meio. Esta leitura contemporânea da psicofisica pode ser melhor apreendida a partir de algumas colocações explícitas de S.S. Stevens. Stevens (1975, p.51-52) propôs que a sensação com a qual a ciência lida é o tipo de reação que se presta ao escrutínio público; que as reações que interessam são aquelas produzidas por organismos em circunstâncias que fazem com que respondam a estímulos; e que a adesão a este princípio nos permite extrair a ordem quantitativa de nossos sistemas sensoriais quando eles respondem a configurações de estímulo do ambiente. Vale a pena ressaltar a acepção proposta por Stevens acerca da noção de subjetividade em psicofisica. Eie afirma que a diferença principal entre os construtos da Física e os da psicofisica é que os construtos psicofísicos dizem respeito a pessoas, chamados observadores ou sujeitos, e são portanto ditos subjetivos. Se lembrarmos que "subjetivo" pertence a "sujeito", em particular à resposta de observadores humanos, podemos evitar as questões metafísicas que costumam surgir quando o discurso é centrado na, assim chamada, experiência subjetiva. Embora as afirmações imediatamente anteriores privilegiem o ser humano como objeto de interesse, não há nada nelas que impeça sua ampliação para incluir outros organismos animais.

As principais perguntas psicofísicas

As principais perguntas que a psicofisica procura responder podem ser agrupadas em quatro tipos (Coren, Porac e Ward, 1979). Uma pergunta é: qual a energia mínima que um estímulo deve ter para provocar uma sensação? Por exemplo, podemos querer saber qual o nível de pressão que um som de 500Hz precisa ter para ser minimamente audível quando

apresentado através de fones de ouvido. Dizemos que esta é uma questão de **detecção,** para cuja resposta procuramos definir o **limiar absoluto.** Uma segunda pergunta é: em quanto dois estímulos precisam diferir para provocarem sensações diferentes? Por exemplo, que distância mínima duas pontas cegas apresentadas à pele do dedo indicador precisam ter para serem percebidas como duas, ao invés de uma? Dizemos que esta é uma questão de **discriminação,** para cuja resposta procuramos definir o **limiar diferencial.** Uma terceira pergunta é: o que é o estímulo?. Um exemplo é a correta leitura de letras da carteia de Snellen, solicitada pelo oftalmologista na avaliação clínica de acuidade visual. Dizemos que esta é uma questão de **reconhecimento,** para cuja resposta procuramos definir o **limiar de reconhecimento** ou a **função de identificação.** Uma quarta pergunta é: como cresce a magnitude da sensação com a magnitude do estímulo? Por exemplo, como cresce a sensação de frio ou desconforto com a queda na temperatura irradiada em uma determinada área da pele? Dizemos que esta é uma questão de **escalonamento,** para cuja resposta procuramos definir uma **função de magnitude psicofisica.** Uma das proposições da psicofisica é encontrar os princípios ou leis gerais subjacentes ao comportamento dos organismos nos tipos de situações acima descritos e, obviamente, as formas metodologicamente apropriadas de se chegar a estes enunciados. O presente capítulo não faz jus à complexidade teórica e metodológica da psicofisica. O tratamento mais aprofundado de vários aspectos da psicofisica pode ser encontrado nas obras relacionadas no final do capítulo.

Leis psicofísicas básicas

Lei de Weber

Voltemos ao experimento precursor sobre discriminação de pesos levantados. Weber observou que pesos pequenos precisam diferir em pouco para serem percebidos como diferentes e que pesos grandes precisam diferir em muito para serem percebidos como diferentes. Dito de outra forma, isto quer dizer que o valor do limiar diferencial para um determinado estímulo varia em função do valor do estímulo padrão com o qual está sendo comparado. Assim sendo, o limiar diferencial não é uma forma econômica de expressar a capacidade discriminativa de um observador em uma dada tarefa sensorial. É conveniente recorrer à expressão da proporcionalidade entre o valor do estímulo padrão e o limiar diferencial. Esta proporcionalidade pode ser matematicamente expressa da forma abaixo, como a lei de Weber:

$$c = (S_2 - S_1)/S_1 \,,$$

onde **c** é a fração ou constante de Weber, S, é o estímulo padrão e S_2 é um estímulo a ser discriminado de S_1.

A lei de Weber tem sido testada para vários tipos de estímulos, mostrando-se aplicável em um conjunto grande de situações. Com seu uso podemos fazer afirmações gerais importantes, como, por exemplo, acerca da comparação entre espécies quanto à discriminabilidade de estímulos em uma dada modalidade sensorial, discriminabilidade relativa dos vários sistemas sensoriais para um mesmo tipo de estímulo em uma dada espécie. A lei tem uma limitação igualmente bem documentada, que é a sua não aplicabilidade para estímulos de magnitude muito pequena. Os dados tipicamente mostram que quando os estímulos têm magnitude muito pequena, c é relativamente grande e tende a diminuir com o aumento na magnitude do estímulo até atingir um valor a partir do qual passa-se a verificar a constância expressa na lei. Em algumas condições específicas, a constância também não se verifica para valores muito elevados de estímulo.

Lei de Fechner

Baseado nos pressupostos de que a diferença apenas perceptível (DAP) é a unidade de sensação e que diferentes DAPs correspondem a um mesmo incremento psicológico em magnitude sensorial, Fechner propôs constituir uma escala de magnitude sensorial somando DAPs a partir do limiar absoluto. Fechner pressupôs também a validade da lei de Weber e, a partir dela, derivou a fórmula que ficou conhecida como a lei de Fechner, abaixo representada:

$$P = k \log S$$

onde P é a magnitude da sensação, k é uma constante de multiplicação cujo valor depende da modalidade e da dimensão sensorial e S é a magnitude (intensidade) do estímulo acima do limiar. Como mencionado anteriormente, os dados empíricos de discriminação mostram limitações nas condições em que se verifica a lei de Weber. Assim sendo, a lei de Fechner é válida apenas nas condições em que se sustentam os pressupostos nos quais está ancorada, um deles sendo a própria lei de Weber.

Lei de Ekman

Um problema de interpretação de dados sobre magnitude de sensação é que as escalas produzidas por escalonamento de razão e as produzidas por escalonamento de discriminação diferem sistematicamente em expoente. Gosta Ekman, trabalhando na Universidade de Estocolmo, procurou uma explicação para esta diferença e propôs, no final da década de 50, que o tamanho psicológico do DAP é uma função linear de magnitude sensorial. A lei de Ekman afirma que:

$$DY = k_Y (Y + a),$$

onde DY é uma mudança em magnitude apenas detectável, Y é o valor de início da magnitude de sensação e k_Y e a são constantes. Em essência, a

lei de Ekman é uma afirmação de que a variabilidade em magnitude sensorial é proporcional ao valor médio de magnitude sensorial. Um aspecto relevante da lei de Ekman é que ela permite unificar dados de experimentos usando métodos de detecção de sinal, discriminação, estimativa de magnitude e julgamento de comparação de pares, cujas características de variabilidade são compatíveis com as descritas na lei. Como digressão, é interessante notar a observação de Gescheider de que, já em 1874, Brentano havia proposto que a relatividade da variabilidade aplica-se não só ao domínio físico como psicológico, mas essa idéia só foi retomada e avançada 80 anos depois. Este é um exemplo específico de um fato corriqueiro na construção da Ciência — de que a evolução do pensamento científico e a acumulação do conhecimento às vezes seguem caminhos bastante tortuosos.

Lei de Stevens

Stevens, em contraposição direta a Fechner, propôs na década de 50 que a razões iguais de estímulo correspondem razões iguais de sensações. Descrita matematicamente, esta relação pode ser reproduzida por:

$$P = k\,S^n,$$

onde P é a magnitude da sensação, k é uma constante de proporcionalidade, S é a magnitude do estímulo e n é a variável potência.

O parâmetro mais importante e que tem sido o foco de muitos experimentos psicofísicos é o expoente n. Quando $n = 1$, a representação gráfica da função é uma reta. Isto corresponde a uma situação psicofisica em que a magnitude da sensação cresce de forma proporcionalmente equivalente, linearmente, com a magnitude do estímulo (vide Figura 2-1). Este fenômeno ocorre, por exemplo, com julgamento de comprimento de linha. Quando $n < 1$, a representação gráfica da função é uma curva com crescimento em aceleração negativa, cuja assíntota tende a uma paralela ao eixo da abcissa. Isto corresponde a uma situação psicofisica em que a magnitude da sensação cresce de forma gradativamente menor conforme aumenta a magnitude do estímulo. Este comportamento da função é o mais freqüentemente encontrado. Dois bons exemplos são o crescimento de brilho em função de intensidade de flashes rápidos e o crescimento de sonoridade em função de pressão acústica. Quando $n > 1$, a representação da função é uma curva de aceleração positiva, com assíntota tendendo a uma paralela ao eixo da ordenada. Isto corresponde a uma situação psicofisica em que a magnitude da sensação cresce de forma gradativamente maior com o aumento na magnitude do estímulo. Este fenômeno ocorre, por exemplo, com choque elétrico em função de corrente aplicada aos dedos.
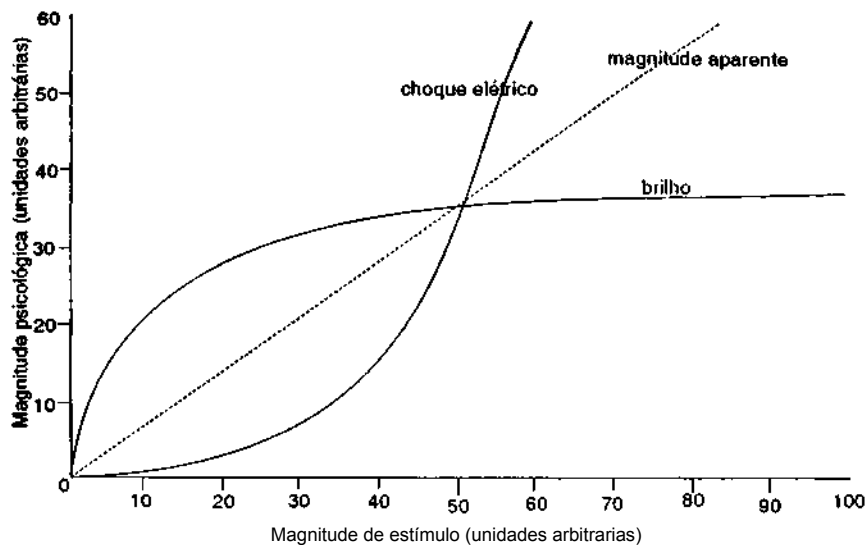
Fig. 2-1 — Representação estilizada dos dados típicos de experimentos relacionando magnitude aparente a magnitude de estímulo para três condições diferentes de estímulo. A magnitude aparente está expressa em unidades arbitrárias como função de magnitude de estímulo, em unidades arbitrárias para choque elétrico, comprimento de linha e brilho. (Adaptado de Stevens, 1975, p.16)

Independentemente da propriedade teórica desta função e de sua eventual superioridade em relação à lei de Fechner, a equação anteriormente descrita apresenta a conveniência de, ao ser logaritmicamente tratada, se transformar na equação da reta:

$$\log P = \log k + n \log S,$$

onde **log** k é o intercepto no eixo **log** P e **n** é a inclinação da reta (vide Figura 2-2).

Nesta forma, a função psicofisica sempre pode ser descrita por uma reta, quando se piota log P em função de log S. Quando n = 1 obter-se-á uma reta de inclinação 1; quando n > 1 obter-se-á uma reta com inclinação superior a 1; e quando n < 1, obter-se-á uma reta com inclinação inferior a 1. Os valores de n e de k podem ser obtidos através do método dos quadrados mínimos.

Rumo a uma lei psicofisica unificada

A comparação entre as leis de Fechner e de Stevens tem ocupado os teóricos mesmo décadas depois do início da controvérsia originalmente
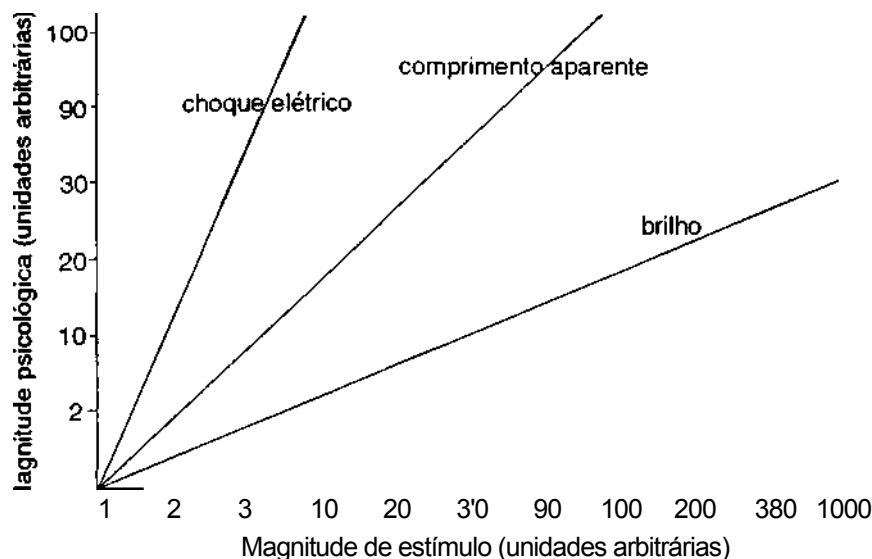
Fig. 2-2 — Representação estilizada dos mesmos dados da Fig. 2-1, logaritmicamente transformados. Observe-se que as três funções são retas, que a curva de aceleração positiva se transformou em uma reta de inclinação maior que 1 e que a curva de aceleração negativa se transformou em uma reta de inclinação menor que 1. (Adaptado de Stevens, 1975, p.17)

colocada por Stevens. O leitor interessado numa visão panorâmica e contemporânea da controvérsia, bem como de uma possível saída para a controvérsia, pode consultar uma publicação no volume 12 do *Behavioral and Brain Sciences* (Krueger, 1989), que reúne um artigo central de Lester E. Krueger com comentários de 36 outros autores. Neste trabalho, Krueger propõe que a descrição de uma lei psicofisica geral exigiria que os fechnerianos abandonassem os pressupostos da validade da lei de Weber e de que o DAP tem a mesma magnitude subjetiva independentemente de modalidade sensorial; e que os stevensonianos abandonassem o pressuposto de que a escala não ajustada de magnitude é uma medida direta de magnitude sensorial. É ainda cedo para avaliar o impacto desta proposta sobre a consolidação da teoria psicofisica.

PSICOFISICA CLASSICA

Fechner criou três métodos psicofísicos, os chamados métodos psicofísicos clássicos, que, com pequenas modificações, continuam sendo considerados métodos centrais na psicofisica.

Método dos estímulos constantes

Neste método, apresenta-se ão observador um conjunto de estímulos cujos valores são adequadamente espaçados, para se obter uma função psicofisica em que a probabilidade de resposta ao longo do contínuo de valores de estímulo varie de 0 a 1. Para a dimensão de estímulo escolhida, cada valor de estímulo é apresentado um mesmo número de vezes, mas em ordem aleatória. Na versão mais simples de sua utilização, o observador é solicitado a julgar se um estímulo está presente ou não; os resultados são plotados em termos de p(SIM) como função de intensidade de estímulo. A função resultante aproxima-se de uma sigmòide e descreve como cresce a sensação acerca do estímulo em função de sua magnitude. Desta função é possível obter-se, como limiar absoluto para a dimensão de estímulo em estudo, aquele valor de estímulo que corresponde à probabilidade 0,5 de respostas SIM. Este valor de probabilidade é visto como aquele em que existe incerteza máxima acerca da ocorrência ou não do estímulo em estudo. É costumeira uma variação de regra de apresentação de estímulo em que a seqüência de estímulos é pseudo-aleatòria, para impedir a ocorrência de séries longas de apresentações de um mesmo valor de estímulo.

Quando o método dos estímulos constantes é usado para a obtenção de limiares diferenciais, o julgamento solicitado ao observador passa a ser, na versão mais simples do método, se um estímulo S_1 é maior ou menor que um segundo estímulo S_2 É variado o tamanho da diferença entre S_1 e S_2, e os valores do par de estímulo ao qual se associa a probabilidade 0,75 de respostas MAIOR QUE ou MENOR QUE são então usados para calcular o limiar diferencial para incremento ou decremento, respectivamente, a partir do valor padrão de S. O estudo de limiares diferenciais exige que o método contenha provisão para lidar com a comparação entre dois estímulos. Assim, por exemplo, em audição, a apresentação simultânea de dois tons produz efeitos de interação que não são triviais; em visão e tato, a apresentação simultânea de dois estímulos pode resultar em estimulação de áreas receptoras da retina ou da pele funcionalmente não comparáveis. A apresentação sucessiva de estímulos sensoriais também tende a produzir efeitos de ordem — outras variáveis permanecendo constantes, o estímulo apresentado em segundo lugar é percebido como de maior magnitude. O conhecimento detalhado do sistema sensorial específico com o qual se está trabalhando é necessário para que sejam incorporados os controles apropriados.

Método dos limites

Neste método o pesquisador predefine urna amplitude de variação de estímulos a serem investigados, bem como os valores específicos de estímulo a serem apresentados ao observador, da mesma forma como o faria no método dos estímulos constantes. Os estímulos são apresen-

tados em séries nas quais os valores de estímulo são de magnitude crescente, tentativa a tentativa (série ascendente), ou de magnitude decrescente, tentativa a tentativa (série descendente). Em cada série, o valor de estímulo com o qual a série se inicia é claramente subliminar (série ascendente) ou supraliminar (série descendente). A série é encerrada quando o observador muda a sua resposta ao estímulo, por exemplo: de NÃO (a tentativa não contém estímulo) para SIM (a tentativa contém estímulo) em uma série ascendente; ou de SIM para NAO em uma série descendente. No jargão psicofisico, estes pontos de mudança são chamados transições. Tipicamente o procedimento contém controles para assegurar que séries de um mesmo tipo não sejam iniciadas com o mesmo valor inicial e que uma mesma sessão de coleta de dados contenha tanto séries ascendentes como descendentes, para impedir ou contrabalançar estratégias de resposta que estejam sob o controle de variáveis outras que magnitude de estímulo. Limiares locais são definidos a partir dos pontos de transição de resposta nas séries individuais, e um limiar médio é calculado a partir da média dos limiares locais.

O método também se aplica ao estudo de limiares diferenciais, circunstância em que se entende como série ascendente aquela em que a diferença entre os estímulos é crescente e série descendente como aquela em que a diferença entre os estímulos tende a zero. As preocupações anteriormente colocadas com as conseqüências da opção por apresentação simultânea ou sucessiva de estímulos também se aplicam a este método. O método dos limites deu origem a três variações: o método da escada, o método do rastreamento (do original inglês *tracking)* e o método de escolha forçada. Apenas o primeiro será aqui descrito; o segundo será omitido porque partilha algumas características com o método dos ajustes, o qual está descrito nesta mesma seção; o terceiro, pelo seu amplo uso associado à teoria de detecção de sinal, será descrito na próxima seção.

Método da escada

Este método avança em relação ao método dos limites a característica de economia no número de tentativas para a obtenção de medidas de limiar. Difere do método dos limites no seguinte aspecto: encerrada a primeira série de estímulos, o valor de estímulo com que se iniciará a série complementar seguinte será o valor com que se encerrou a série anterior. A conseqüência disto é que os valores de estímulo efetivamente apresentados estão sempre muito próximos do limiar. Obtém-se um grande número de estimativas de limiar em curto período de tempo. Da mesma forma como no método dos limites, limiares locais são definidos a partir dos valores de estímulo correspondentes a transições de resposta, e um limiar médio é obtido a partir da média dos limiares locais. Procura-se obter um número de transições que seja um compromisso entre o neces-

sário para o uso confortável de estatística descritiva e a possibilidade de várias aferições de limiar para a construção de uma função. Este método, da mesma forma que o método dos limites, é encontrado numa série de variações no que diz respeito a quatro características: (a) os critérios usados para determinar mudanças no nível do estímulo; (b) o valor, ou o tamanho do intervalo de valor, em que o estímulo é aumentado ou diminuído de tentativa a tentativa; (c) o critério usado para encerrar a escada; e (d) o algoritmo usado para calcular o limiar a partir dos dados.

Método dos ajustes

Este é o terceiro método psicofisico concebido por Fechner. Foi originalmente concebido para a mensuração de limiares diferenciais, mas pode ser usado também para o estabelecimento de limiares absolutos. Na versão para limiares absolutos, o estímulo é inicialmente apresentado em um valor bem acima do limiar e pede-se ao sujeito para ajustar o nível do estímulo até que ele desapareça. Em seguida, o estímulo pode ser apresentado bem abaixo do limiar e pede-se ao sujeito para ajustar o nível do estímulo até ele ser apenas perceptível. Tipicamente, o estímulo varia continuamente. Uma série de tentativas em que o estímulo cresce ou decresce é apresentada. O limiar é calculado como a média dos diversos valores de ajuste.

A versão para limiares diferenciais tem também o nome de método do erro médio. Nele, apresenta-se ao observador um estímulo padrão e um segundo estímulo, o qual ele é instruído a ajustar para que pareça igual ao padrão. Como o uso típico do método é poder descrever a discrepância entre o valor de estímulo padrão e o valor de ajuste, o tratamento de dados é voltado para este aspecto. A média da distribuição dos ajustes é considerada o valor de **igualdade subjetiva;** o **erro constante** é a diferença entre os valores do estímulo padrão e os da igualdade subjetiva; o desvio padrão (ou outra medida de dispersão) é o **limiar diferencial.**

Uma limitação do método dos ajustes é que a variação contínua do estímulo torna o resultado pouco preciso no caso de aferição de limiar. Uma outra limitação é que tempo de exposição e taxa de mudança por si sós podem ser variáveis relevantes na sensação que um estímulo produz, e o experimentador perde controle sobre estas variáveis se o procedimento permite que o observador manipule estes aspectos do estímulo.

PSICOFÌSICA MODERNA

Stevens, em um trabalho volumoso que se iniciou na década de 30, questionou a formulação básica da lei de Fechner, de que a razões iguais de estímulo correspondem diferenças iguais de sensações, e contestou

diretamente o pressuposto fechneriano de que as sensações não podem ser medidas diretamente. Propôs uma nova lei de acordo com a qual a razões iguais de estímulos correspondem iguais razões de sensação, a lei da potência, posteriormente também chamada de lei de Stevens, e demonstrou que é possível obter escalas confiáveis relacionando atributos físicos a atributos psicológicos, perguntando-se diretamente ao observador acerca de propriedades dos estímulos. Criou os chamados métodos de escalonamento direto, dos quais destacaremos quatro neste capítulo.

Os métodos de escalonamento e as estatísticas a eles associadas para produzir diferentes escalas sensoriais têm por trás a caracterização do grau de correspondência entre o sistema de números e as propriedades dos objetos ou eventos em estudo. De acordo com estas propriedades, as escalas de medida são agrupadas em quatro tipos: escalas nominais, ordinais, de intervalo e de razão. Os métodos de escalonamento podem ser agrupados em duas grandes categorias, situando-se na primeira os métodos para construção de escalas de intervalos. Esta primeira categoria se subdivide em duas famílias de métodos — para a construção de escalas de confusão e de escalas de partição. Na segunda categoria estão os métodos para construção de escalas de razão, sendo apenas estas aqui descritas, dada a preponderância de seu uso na psicofisica contemporânea. O leitor pode consultar Stevens (1975) ou Gescheider (1976) para uma boa descrição de outras escalas.

Iniciativas precursoras de uso de escalas de razão remontam a Merkel em 1888 e a Fullerton e Cattell em 1892, mas foi só na década de 30, quando os engenheiros acústicos começaram a se preocupar com uma descrição numérica de sonoridade *(loudness),* que os psicólogos passaram a investir esforços neste tipo de escala. Os quatro principais métodos para produção de escalas de razão serão descritos a seguir.

Método de produção de razão

Neste método, também chamado de fracionamento, solicita-se ao observador que observe um estímulo padrão e ajuste um segundo estímulo, variável, até que este pareça um certo valor de razão do primeiro. Por exemplo, fornece-se um tom puro de uma certa freqüência a uma certa intensidade e solicita-se ao sujeito que ajuste a intensidade de um segundo tom de mesma freqüência, até que este pareça ter a metade da sonoridade do primeiro tom. A tarefa é repetida para vários valores de intensidade de estímulo. A tarefa pode também ser de razões que se expressem por múltiplos, como, por exemplo, achar o valor que reflita o dobro, o triplo de sonoridade. Recomenda-se que a descrição de uma função de magnitude seja baseada num experimento que inclua os dois tipos de tarefa. A forma como os estímulos são apresentados costuma ser

especificada de acordo com um dos métodos da psicofisica clássica anteriormente descritos. Uma variação do método é fazer com que o estímulo variável assuma valores discretos, espaçados a pequenos intervalos de intensidade, em vez da variação contínua, cuja desvantagem de quantificação já foi anteriormente exposta. O método de produção de razão já foi utilizado para a obtenção de funções de magnitude para várias modalidades e dimensões de estímulo.

Método de estimativa de razão

Neste método, pede-se ao sujeito para, dados dois estímulos que se diferenciam em intensidade por um valor definido pelo experimentador, estimar a razão entre eles. Aplicam-se as mesmas considerações feitas na descrição imediatamente acima para a obtenção de uma função de magnitude. Os métodos de produção e de estimativa de razão podem ser usados para validação recíproca. Gescheider (1976) alerta para o fato de que estes dois métodos não são muito poderosos, porque são muito influenciáveis por variáveis de contexto, como, por exemplo, a própria amplitude dos valores de estímulo usados no experimento.

Método de estimativa de magnitude

A genialidade deste método está na sua surpreendente simplicidade. Pede-se ao observador para atribuir números a variações quantitativas em um atributo de um estímulo. Em uma das variantes do método, dá-se ao observador um estímulo padrão ao qual se atribui um valor de medida, do tipo: se o peso deste cilindro tem o valor de 100, em quanto você estima o peso de uma série de cilindros que lhe serão apresentados? Stevens recomenda que o estímulo padrão corresponda a um valor intermediário, dentre aqueles que serão julgados. Em outra variante, o sujeito é livre para atribuir valores, e a instrução se limita a orientar o sujeito no sentido de que os valores atribuídos reflitam seu julgamento da magnitude relativa dos estímulos. Nesta segunda variante, a comparação de dados entre diferentes observadores exige que se trabalhe com média geométrica ou que se use um procedimento preliminar de normatização dos valores de resposta para os diferentes observadores. Como a tarefa é tipicamente considerada muito fácil pelo observador, a coleta de dados é muito eficiente, oque torna o método recomendado em experimentos que exigem a obtenção de muitos dados. O leitor pode consultar Gescheider (1988) para uma análise de modelos explicativos das características dos dados obtidos por estimativa de magnitude.

Método de produção de magnitude

Este método é a recíproca do anterior. Dá-se ao observador um estímulo padrão num dado valor especificado pelo experimentador e pede-se ao observador para ajustar um segundo estímulo até que o valor deste

seja um determinado múltiplo ou fração do padrão. Novamente, a validade da escala de magnitude é julgada pela boa aproximação entre as escalas obtidas por produção e por estimativa de magnitude.

Comparando-se os resultados obtidos com os métodos de estimativa e de produção de magnitude, observa-se um viés típico na inclinação de uma função em relação a outra, descrito como efeito de regressão à média, com a seguinte característica: na tarefa de estimativa, o observador evita fazer estimativas muito baixas ou muito altas, resultando em uma inclinação menor da função; já na tarefa de produção, o observador evita ajustar o estímulo em valores muito baixos ou muito altos, resultando em uma maior inclinação da função. O balanceamento destes dois vieses complementares pode ser feito através do **método de balanceamento numérico de magnitude,** que consiste na extração da média geométrica das duas funções.

Método de equiparação intermodal

Diferentemente dos métodos psicofísicos acima apresentados, nos quais se busca descrever algum tipo de função R = f (S), este método é voltado para a obtenção de funções do tipo S = f (S). Rigorosamente falando, não é um dos métodos básicos, mas suas peculiaridades merecem uma atenção especial. Neste método, o observador é solicitado a fazer comparações diretas entre dois contínuos sensoriais. Uma tarefa típica seria solicitar ao observador apertar um dinamòmetro com força equiparável à intensidade aparente de um som. Este método foi desenvolvido com uma preocupação teórica acerca da lei potência. Stevens argumenta que, se a estimativa numérica que um observador faz quando **é** solicitado a julgar a magnitude de sua sensação reflete uma propriedade fundamental do sistema sensorial, então esta mesma propriedade deveria nortear seu comportamento quando ele equipara uma sensação a outra. Assim, conhecendo-se a função potência para cada sensação, a equiparação deveria ser uma outra função potência cujo expoente é a razão dos dois expoentes originais. A aproximação entre o expoente predito e o empiricamente obtido na tarefa de equiparação pode, então, ser usada como indicador de validade da escala psicofisica. As implicações teóricas da equiparação intermodal continuam sendo objeto de pesquisa. O leitor pode recorrer a Luce (1990) para um tratamento avançado contemporâneo do assunto.

DETECÇÃO DE SINAL

Tanner e Swets propuseram em 1954 que a teoria de decisão estatística e certos conceitos sobre dispositivos eletrônicos de detecção de sinal poderiam ser usados para construir um modelo que se aproximasse de como as pessoas se comportam em situações de detecção nas quais

existem expectativas e conseqüências. O modelo ficou conhecido como a teoria de detecção de sinal e está descrito no livro *A Signal Detection Theory and Psychophysics,* de D.M. Green e J.A. Swets, publicado em 1966 e considerado a referência básica principal para a teoria. A descrição que se segue é baseada em Gescheider (1976).

A teoria de detecção de sinal se contrapõe à psicofisica clássica ao rejeitar a relevância do conceito de limiar e ao propor que o observador não recebe passivamente o estímulo, mas, muito ao contrário, está ativamente engajado no processo de decidir acerca da presença de um estímulo na situação de observação. Assim, a teoria e o método estão voltados para a análise do comportamento de tomada de decisão do observador. A teoria contém algumas proposições centrais: a) existe atividade sensorial, mesmo na ausência de estímulo específico; b) um sinal sempre ocorre na presença de ruído; c) o nível de ruído varia de momento a momento; d) o observador define (conscientemente ou não) acerca de um **critério** ou ponto de corte para um nível de sensação acima do qual responde SIM (existe um estímulo, ou sinal, como preferem os usuários da teoria) e abaixo do qual, NÃO; e) sensibilidade (detectabilidade ou discriminabilidade, dependendo de detalhes da tarefa psicofisica específica) é a quantidade média de sensação gerada por um sinal quando comparado com a quantidade média de sensação gerada por ruído apenas. Duas medidas centrais inerentes à teoria são, portanto, critério de resposta (P) e sensibilidade ao estímulo (d'). Além disto, a legitimidade da utilização da teoria pressupõe a verificação dos pressupostos estatísticos de que as distribuições de ruído e de sinal-ruído sejam normais e que as variâncias das respectivas distribuições sejam iguais.

Dentro do arcabouço acima descrito, um experimento psicofisico simples pode ser representado por uma matriz estímulo x resposta 2x2, na qual, para duas possibilidades de sinal (presente ou ausente) contempla-se duas possibilidades de resposta (sim ou não). Uma matriz assim delineada permite verificarem-se quatro possibilidades de resultado de tentativas individuais, sendo dois tipos de acerto e dois tipos de erro. Mais especificamente, se um sinal está presente e a resposta é sim, obtém-se uma **detecção correta;** se um sinal está presente e a resposta é não, uma **omissão;** se um sinal está ausente e a resposta é sim, um **alarme falso;** e se o sinal está ausente e a resposta é não, uma **rejeição correta** do sinal. Como, necessariamente, são complementares as proporções de resposta sim e não para cada uma das possibilidades de sinal, por convenção selecionam-se para estudo as categorias **detecção correta** e **alarme falso,** cujas probabilidades são independentes; este delineamento básico é repetido para probabilidades diferentes de apresentação de sinal. Todo o conjunto pode ainda ser repetido para valores diferentes de sinal (vide Figura 2-3).

Os dados de um experimento assim planejado e executado podem ser graficamente apresentados como curvas de isossensibilidade quando
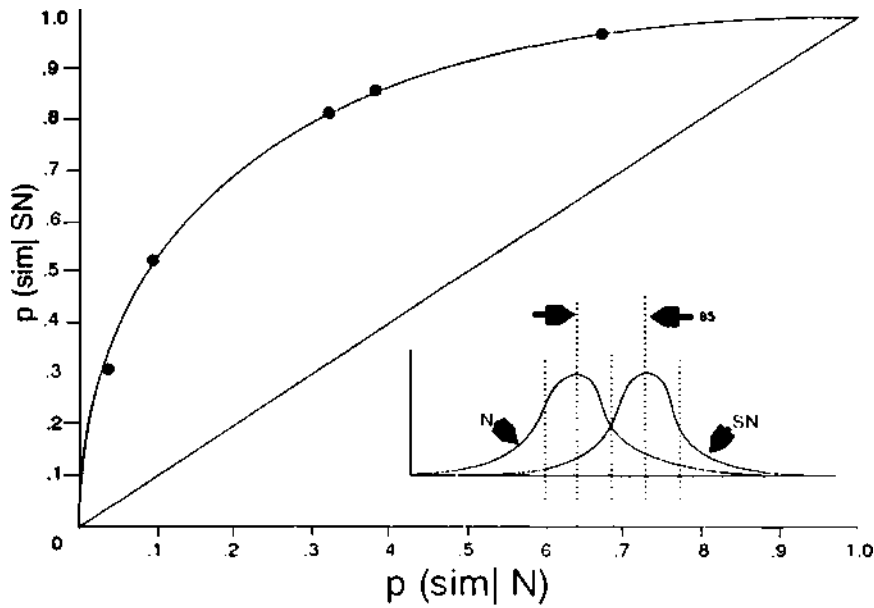
Fig. 2-3 — Representação estilizada de uma curva de isossensibilidade a partir de um experimento de detecção de sinal, em que cada ponto representa os dados sob uma condição de probabilidade de ocorrência do sinal. A probabilidade de correta detecção está expressa na ordenada e a probabilidade associada de alarme falso está expressa na abcissa. (Adaptado de Stevens, 1976, p.74)

a proporção de detecções corretas é plotada na ordenada, a proporção de alarmes falsos é plotada na abcissa e cada curva representa o comportamento de detecção para um único valor de estímulo associado a várias probabilidades de apresentação de sinal. Curvas diferentes descrevem a sensibilidade do observador a valores diferentes de estímulo. Com este tipo de apresentação de dados, a curvatura da função, ou o grau de seu afastamento de uma reta de inclinação 1, indica a sensibilidade do observador ao estímulo. Para testar os pressupostos de normalidade de distribuições e de igualdade de variância, a curva de isossensibilidade deve ser transformada em uma reta, o que se faz transformando as proporções em escores Z. Feita esta transformação, se a distribuição for normal, a função será uma reta; se as variâncias forem iguais, a inclinação será 1 — e os valores de d' serão os valores de intercepto. A reta é costumeiramente descrita a partir do método dos quadrados mínimos.

A solução gráfica não é, porém, necessária para descrever os resultados do experimento. O valor de d' pode ser diretamente calculado a partir

da proporção de detecções corretas e de alarmes falsos através da fórmula:

$$d' = [\sigma_{SR} / \sigma_{R}][Z_{\text{detecções corretas}}] - Z_{\text{alarmes falsos}}$$

O procedimento típico comporta a manipulação explícita de variáveis motivacionais, pois é possível substituir a manipulação de freqüência relativa do sinal por manipulação de tipo e magnitude de conseqüência para cada uma das quatro possibilidades de resposta. Desta forma pode-se verificar a influência de variáveis não estritamente sensoriais sobre o padrão de respostas do observador.

Como anteriormente mencionado, a teoria tem um referencial estatístico. Embutidos em um método geral acima descrito, pode-se descrever três procedimentos associados à teoria de detecção de sinal: o procedimento **sim-não,** o procedimento de **escolha forçada** e o procedimento de **julgamento** *(rating)* **de confiança.** É importante apreciar que estes procedimentos são delineados para permitir a obtenção de dados que possam, por sua vez, ser convertidos nos construtos teóricos de sensibilidade, de critério, de distribuição de variância e de forma de distribuição, e permitir o teste da teoria comparando-se os valores dos construtos preditos na teoria com aqueles derivados dos dados empíricos de proporção de respostas.

Procedimento sim-não

O experimento é organizado em tentativas de dois tipos. Em algumas tentativas o sinal é apresentado (mais rigorosamente, tentativas contendo sinal e ruído) e em outras tentativas o sinal não é apresentado (mais rigorosamente, tentativas contendo apenas ruído). O observador é informado da matriz de pagamento (ou, em linguagem coloquial, do que tem a ganhar e perder com acertos e erros, respectivamente). Sua tarefa é simplesmente julgar, através de uma resposta simples do tipo sim/não, se um sinal está ou não presente na tentativa Em cada sessão experimental vigora uma condição de matriz de pagamento e de probabilidade sinal/ruído.

Poder traçar uma curva de isossensibilidade exige que o experimento inclua várias condições de probabilidade sinal/ruído, o que muitas vezes não é feito, pois isto exige várias sessões experimentais e um número muito grande de tentativas. Nestas circunstâncias não existem condições de se testar os pressupostos de distribuição normal e variâncias iguais, tendo-se que usar um tratamento estatístico que dispense a verificação destes pressupostos.

Procedimento de escolha forçada

Neste procedimento, o observador é exposto a tentativas que contêm dois ou mais intervalos de observação e deve indicar qual deles contém o sinal. Pressupõe-se que, na ausência de viés de resposta em

relação a um ou outro intervalo de observação, o observador escolha o intervalo que contém o estímulo de maior magnitude. Na prática, o planejamento das tentativas deve incluir controles para impedir o desenvolvimento de vieses de resposta, como, por exemplo, viés de posição ou de ordem. A própria proporção de respostas corretas é considerada medida direta de sensibilidade.

Procedimento de julgamento de confiança

Em cada tentativa o observador deve indicar quão seguro está da presença do sinal, em uma escala de grau de confiança apresentada pelo experimentador. Pressupõe-se que o dado de estimativa é medida de critério, o que representa uma economia em coleta de dados. Para fins de análise de dados, as respostas são desdobradas por valor de grau de confiança, e com os dados de uma sessão experimental consegue-se obter uma função de isossensibilidade. Este procedimento é considerado vantajoso em relação aos dois anteriormente descritos, não só pela rapidez como pela conveniência para testes de normalidade de distribuição e igualdade de variância.

PSICOFÌSICA ANIMAL

Estudos precursores em psicofisica animal apareceram em 1935, quando a equipe de Culler e Brodgen desenvolveu um método cuidadoso e sensível para obtenção de limiares auditivos no cachorro, usando esquiva. Ao se apontar um marco para o surgimento formal da psicofisica animal, seria justo destacar um outro trabalho, pelo seu impacto na comunidade científica — uma publicação de Donald Blough, em 1958, na qual ele propõe um procedimento para a obtenção de uma função psicofisica de adaptação visual no pombo. Na década de sessenta, vários laboratórios se dedicaram ao desenvolvimento de metodologias de pesquisa apropriadas ao estudo das sensações em várias modalidades e em várias espécies animais. Estas múltiplas iniciativas foram consolidadas no livro *Animal Psychophysics,* organizado por William C. Stebbins e publicado em 1970. Outro trabalho importante de consolidação dos métodos da psicofisica animal é um texto de autoria do casal Donald e Patrícia Blough, publicado em 1977.

Não se pode dizer que a psicofisica animal tenha um referencial teórico próprio, embora ela possa causar um certo desconforto aos puristas teóricos, que vêem a psicofisica como um empreendimento para destrinchar questões referentes à relação mente-corpo. Porém ela, clara e explicitamente, faz uso do conhecimento científico sobre o comportamento animal desenvolvido tanto na perspectiva etologica quanto de análise experimental do comportamento. A psicofisica animal se propõe a desenvolver procedimentos específicos que permitam a utilização de metodolo-

gias básicas independentemente desenvolvidas, como as já descritas neste capítulo, quer para assegurar o estudo de determinadas espécies, quer para permitir estudos comparativos entre espécies. As questões que lhe interessam são as mesmas leis e princípios gerais anteriormente descritos, as teorias, leis e princípios próprios dos diversos sistemas sensoriais e a verificação do seu grau de generalidade no reino animal. A seguir serão descritos, por categoria de problema, os procedimentos recomendados para adaptar os mais variados métodos psicofísicos ao sujeito animal.

Aspectos motivacionais

A impossibilidade de usar instruções verbais, muito ao gosto dos estudiosos da psicofisica humana, leva ao uso de regras de contingência. O estabelecimento destas regras não é trivial. Tomemos, por exemplo, o caso de especificação de limiar absoluto. Se é razoável que uma resposta de detecção na ausência de estímulo seja considerada uma resposta errada e, portanto, deva não ser reforçada ou até mesmo punida, não é razoável planejar que a ausência de resposta na presença de um estímulo fraco seja considerada erro, pois é uma incógnita a magnitude de estímulo necessária para que ele seja estímulo efetivo. O uso de reforçamento positivo, ao mesmo tempo que é eficaz em manter o animal engajado na tarefa, impõe limites ao número de tentativas contidas em uma sessão experimental, em virtude de saciação do reforçador, o que leva ao cuidado de que o procedimento otimize a razão tentativas úteis/tentativas totais.

Resposta de observação

A baixa variância dos dados psicofísicos depende de que o animal esteja atento ao estímulo sensorial e adequadamente posicionado, de modo que o estímulo entre em contato com o órgão receptor de forma especificável e constante. Tipicamente uma tentativa é iniciada com a apresentação de um estímulo que sinaliza a ocasião para a emissão de uma resposta de observação. Isto significa, por exemplo, no caso de estímulos visuais, que o procedimento seja tal que o estímulo sensorial só será apresentado se o animal estiver com a cabeça posicionada de uma maneira determinada, para assegurar que os olhos estejam a uma distância e orientação especificadas em relação ao estímulo. Em um procedimento típico, a emissão de uma resposta de orientação é condição para que se inicie o intervalo que pode ou não conter o estímulo sensorial. A correta emissão da resposta de observação é reforçada com um estímulo arbitrário que também exerce a função de sinalizar o início de um intervalo de apresentação de estímulo. A topografia desta resposta depende de variáveis próprias do experimento, como a espécie animal, a modalidade e a dimensão de estímulo em estudo.

Controle de estímulo da resposta de detecção

O procedimento precisa incluir salvaguardas para que não existam outras variáveis, além do próprio estímulo sensorial em estudo, exercendo o papel de sinalizadores para a ocorrência da resposta de detecção. A própria existência de uma sinalização para a ocorrência do estímulo sensorial pode propiciar a ocorrência de respostas indevidas de detecção. Um animal pode estimar o intervalo de tempo entre eventos relevantes e convenientemente espaçar suas respostas de forma a maximizar a obtenção de reforços. Embora esta afirmação costume soar como um exagero para aqueles não familiarizados com o estudo de aprendizagem animal, ela está fartamente documentada na literatura especializada. Salvaguardas típicas para impedir esta possível estratégia de resposta são: tornar variáveis o tempo entre a emissão da resposta de observação e a apresentação do estímulo sensorial e o tempo entre a resposta de detecção e o início da tentativa seguinte; punir respostas de observação e respostas de detecção na ausência de seus respectivos estímulos sinalizadores com um *timeout,* punir respostas de detecção com tempo de reação inferior ao mínimo necessário, considerando o que se conhece sobre tempo mínimo de resposta.

Os controles acima são necessários, mas ainda não suficientes. Por exemplo, um *switch* que aciona um tom pode conter transientes audíveis, fazendo com que a ocorrência do estímulo seja detectável independentemente de sua intensidade e, em conseqüência, o animal seja reforçado pela detecção desses transientes, o que pode levar o pesquisador a concluir pela sua supersensibilidade. Uma forma de controlar isto é introduzir pseudotentativas nas quais tudo ocorre como em uma tentativa regular, exceto que o estímulo é apresentado em uma magnitude aberrantemente distanciada da faixa de sensibilidade conhecida para a espécie. O uso de pseudotentativas propicia uma forma de monitorar a qualidade do desempenho do animal. A porcentagem de respostas a estas tentativas pode ser usada para decidir sobre a confiabilidade dos dados. As respostas a pseudotentativas, dentro de um referencial de teoria de detecção de sinal, podem ser um indicador de critério de resposta. Assim, um animal conservador terá uma baixa proporção de respostas a estas tentativas e um limiar mais alto; um animal menos conservador terá uma proporção mais alta de respostas a estas tentativas e um limiar mais baixo. Um controle sobre este critério de resposta pode ser tentado via manipulação da especificação do *timeout.*

Tempo de reação e escalas psicofísicas

Embora na maioria das tarefas psicofísicas de detecção se solicite ao observador uma resposta simples e arbitrária que é registrada como ocorrendo ou não dentro de um intervalo de tempo no qual a resposta

pode ser dada, é possível medir-se o intervalo entre o início da apresentação do estímulo e a ocorrência da resposta de detecção do estímulo. Este tempo de reação guarda uma relação inversa natural e ordenada com magnitude de estímulo para todas as modalidades testadas, de forma tal que pode ser usado para escalonar estímulos. Se o *manipulandum* e a topografia da resposta reforçada são cuidadosamente adaptados às características da espécie em estudo, é possível obterem-se dados de excelente qualidade. Tempo de reação é um método que nasceu fora da psicofisica; sua invenção é creditada a Franciscus Donders, em torno de 1865, interessado em calcular a velocidade de operações mentais. O método foi incorporado ao laboratório de Wundt e posteriormente aplicado à psicofisica, não com os objetivos originais, mas em função desta sua propriedade de relação natural com magnitude de estímulo. A proposição deste uso para tempo de reação remonta a Cattell, em 1902, mas só a partir da década de 60 tempo de reação começou a ser usado de forma sistemática na psicofisica animal. Vide, por exemplo, a demonstração de seu uso para pesquisar tanto audição quanto visão em macacos, feita por Stebbins e Miller (1964). Este método vem sendo usado para descrever a sensação de brilho e de sonoridade em várias espécies, e os dados se comportam de forma equivalente aos obtidos para observadores humanos, não só em tarefas semelhantes de tempo de reação, mas comparando dados humanos obtidos usando tempo de reação e estimativa de magnitude (Marshall, Brandt, 1980; Pfingst, Hienz, Kimm e Miller, 1975; Reason, 1968). O método tem uma série de variações que não nos é possível detalhar neste capítulo. É possivelmente o método mais usado em psicofisica animal para a pesquisa sobre escalas de sensação.

É importante a apreciação de que as várias adaptações anteriormente descritas, além de permitirem a transposição da barreira da linguagem, asseguram a efetiva observância das regras esperadas de comportamento do observador humano explicitadas através de instrução verbal. Com freqüência, os procedimentos de psicofisica animal podem, com pequenas adaptações, ser usados com seres humanos, permitindo a validação do uso do comportamento animal como modelo do comportamento humano. A combinação da psicofisica animal com eletrofisiologia sensorial, belamente representada pelos trabalhos pioneiros de DeValois e DeValois em visão, permitiram um avanço significativo na psicobiologia dos sistemas sensoriais. Fica aqui apenas o registro, pois este é um outro assunto que mereceria espaço próprio para apresentação.

PSICOFÌSICA DA CRIANÇA

Sea década de 60 foi decisiva para o desenvolvimento de condições para uma psicofisica animal, a década de 80 poderia ser apontada como

aquela em que se concentraram esforços para o desenvolvimento de uma psicofisica da criança. Esta psicofisica tem que dar conta de algumas características básicas do comportamento do organismo em desenvolvimento, como as limitações no repertório com portam ental disponível e alterações neste repertório em função de maturação. Como exemplo da primeira categoria temos a impossibilidade de utilização de instruções e de respostas verbais no infante, um problema que, na sua essência, não é diferente daquele encontrado na psicofisica animal. Como exemplo da segunda categoria temos o fato de que respostas que se mostram adequadas numa faixa de idade deixam de ser utilizáveis numa faixa posterior, criando obstáculos à realização de estudos longitudinais. Não se dispõe ainda de uma sólida metodologia psicofisica da criança, mas há progressos importantes que justificam uma sistematização.

A ordem em que os principais métodos disponíveis serão apresentados segue aproximadamente a ordem de idade aos quais têm se mostrado mais adequados. Como ficará claro, eles são, antes de mais nada, procedimentos que viabilizam a utilização de algum dos métodos psicofísicos propriamente ditos, apresentados mais no início deste capítulo, em conjunção com algum sistema de resposta que se identifique como apropriado para a idade e para o sistema sensorial específico em estudo. Observe-se o recurso sistemático a um sistema de resposta que pode ser caracterizado genericamente como resposta de observação ou de atenção. Nas idades mais tenras predomina o uso de respostas gerais indiferenciadas, e nas idades mais elevadas, o uso de respostas específicas.

Psicofisica do feto

A pesquisa sobre a ontogenèse das sensações tem, obviamente, um interesse especial na descrição de quando se inicia o responder a estímulos nas diferentes modalidades sensoriais. Há várias evidências bem estabelecidas de que as crianças já nascem com competência para detectar uma série de estímulos ambientais. Vide, por exemplo, revisão de Cowart (1981), na área de sentidos químicos, e de Reisman (1987), na área de somatossentidos. Assim sendo, o surgimento dessa competência precisa ser buscado no feto, um empreendimento problemático pela inacessibilidade relativa do observador psicofisico.

Observação extra-uterina do feto

Uma abordagem encontrada na literatura mais antiga é o estudo de fetos não-viáveis e com poucas horas de vida após o aborto. Estes estudos, necessariamente muito primitivos, se concentraram na observação da presença ou não de reflexos (vide revisão de Reisman, 1987) e mostraram respostas bastante precoces a estímulos táteis aplicados à face. Dadas as circunstâncias de coleta de dados, a disponibilidade e interpretação de dados são muito problemáticas.

Observação intra-uterina do feto

Birnholtz e Bernacerraf (1983) mostraram que a ultra-sonografiafetal permite a observação de atividade muscular fetal em resposta a sons apresentados extra-uterinamente. Esta técnica contém uma série de limitações relacionadas a especificação de estímulo sensorial e a repertório comportamental disponível. No presente momento permite informações apenas de natureza qualitativa, embora o caráter não-invasivo e a dissociação de condições patológicas imprimam, em princípio, maior garantia de validade dos dados obtidos. A técnica é promissora especialmente para o estudo da modalidade auditiva, mas ainda não se transformou em uma rotina de pesquisa. Na área dos sentidos químicos, uma possibilidade sugerida em Cowart (1981) é a observação de movimentos de deglutição em resposta a alteração na composição do líquido amniotico, produzida pela introdução de substâncias químicas na cavidade intra-uterina, através de sonda.

Psicofisica do Infante

Procedimentos baseados em atividade geral fisiológica e comportamental

Alterações em medidas como padrão do ciclo de sono, taxa de batimento cardíaco, padrão de respiração, nível de várias substâncias na corrente sangüínea têm sido usados para o estudo das sensações, embora sejam mais encontrados em estudos mais antigos. Schneider, Trehub e Bull (1979) fizeram uma análise comparando os resultados de experimentos sobre limiar auditivo em recém-nascidos usando uma variedade de respostas e encontraram grandes discrepâncias nos dados, argumentando pela sua baixa confiabilidade. Reisman (1987), fazendo uma revisão de pesquisas sobre dor em bebês circuncidados, também encontrou dificuldades na interpretação dos dados com essa categoria de medida. Estudos sensoriais baseados em atividade fisiológica e comportamental gerais, rigorosamente falando, nem se classificam como estudos psicofísicos; sua inclusão aqui é, muito mais, uma chamada de atenção para a precariedade metodológica no estudo da ontogenèse das sensações.

Procedimentos baseados em respostas específicas de orientação

A criança recém-nascida e até aproximadamente de quatro meses de idade tem sido pesquisada com procedimentos principalmente baseados em seu repertório de reflexos incondicionados. A observação da ocorrência de respostas incondicionadas à apresentação de estímulos específicos permite afirmar que a criança detecta o estímulo, mas não tem permitido fazer afirmações confiáveis a respeito de valores mínimos de estímulo necessários para que ocorra uma sensação. Estes reflexos também estão sujeitos a habituação, ou seja, sua apresentação repetida leva ao enfraquecimento e eventual desaparecimento da resposta, fazendo

com que um experimento tenha que se limitar a um pequeno número de observações. Esta limitação não é facilmente contornada pela repetição do experimento a intervalos consideráveis na mesma criança, não só porque o repertório de respostas vai se alterando como porque os próprios sistemas sensoriais estão em desenvolvimento. O experimentador costuma então recorrer à análise de dados de grupo, o que é contrário à tradição psicofisica de privilegiar delineamentos intra-sujeito e análise de dados individuais. Algumas destas respostas específicas vêm sendo estudadas sistematicamente, razão pela qual são identificadas a seguir.

Método do olhar preferencial

Este método, creditado a Fantz e Berlyne, que o desenvolveram simultânea e independentemente, é baseado na tendência observada em crianças pequenas de fixar o olhar em um estímulo visual com padrão e de não fixar o olhar em um estímulo liso, em uma situação de escolha. Tipicamente, são registrados dados referentes à direção da primeira fixação, número de fixações e tempo de fixação em cada estímulo. Interpreta-se que a criança detecta um padrão visual se a ele corresponder maior freqüência e duração de fixação do olhar. Do método do olhar preferencial surgiu uma variante, o método do olhar preferencial de escolha forçada, atribuído a Teller. Neste segundo método, um observador adulto é colocado em uma posição na qual pode observar os olhos da criança, mas não os estímulos sendo apresentados. A tarefa do observador é julgar a localização do padrão com base na observação do olhar da criança, sendo-lhe permitido escolher o(s) aspecto(s) do olhar da criança que julgar mais ¡nformativo(s). Se o julgamento de localização do observador for estatisticamente diferente de desempenho aleatório, interpreta-se que a criança detecta o padrão (Banks, Dannemiller, 1987).

Habituação

Habituação é mais do que um método; é um paradigma a partir do qual uma série de métodos específicos emergiram, adaptados para modalidades diferentes de estímulo. Baseia-se no fato de que, quando um estímulo novo é apresentado, o organismo emite uma resposta de orientação. Com a manutenção do estímulo ou sua apresentação repetida a curtos intervalos de tempo há um decréscimo e eventual desaparecimento da resposta. Sendo o estímulo alterado (novo estímulo apresentado), a resposta de orientação reaparece. Assim, por exemplo, estímulos visuais novos geram uma exploração visual que tende a desaparecer com a manutenção do estímulo, mas reaparece com a troca do estímulo por um novo; estímulos auditivos geram uma resposta de virada de cabeça em direção ao som, a qual tende a desaparecer com manutenção do estímulo. Este paradigma inclui também fenômenos em que a resposta de orientação envolve a inibição de uma atividade em andamento. Por exemplo, a resposta de sucção é interrompida com a apresentação de um som; a ini-

bicão da sucção habitua-se reinstalando-se a sucção com a manutenção do estímulo auditivo, mas a apresentação de estímulo auditivo diferente reinstala a inibição. A interpretação costumeira dos dados é que se ocorrer desabituação, então os estímulos envolvidos são discrimináveis. O leitor pode consultar Dodwell, Humphrey e Muir (1987) para exemplos detalhados do uso do paradigma ao estudo da percepção visual de padrões.

Varredura visual

O método é baseado no fato de que, quando as pessoas exploram visualmente imagens, a varredura da imagem segue padrões característicos (Dodwell, Humphrey e Muir, 1987), e a análise destes padrões de varredura deveriam informar sobre aspectos do estímulo visual aos quais a pessoa responde. Dodwell e colaboradores alertam que a implementação deste método em crianças é bastante complexa, do ponto de vista técnico. Bergamasco (1992) observa que, devido à imaturidade de fotorreceptores em uma faixa precoce de idade, a melhor imagem visual pode não corresponder à visão foveal, o que dificulta a interpretação do dado tipicamente analisado com o uso deste método. Feitas estas ressalvas, o método é útil para o estudo de percepção de forma e de padrão.

Expressões faciais

Uma abordagem adicional tem sido a de caracterizar a relação entre tipos de estímulo e tipos de topografia de resposta. Nesta abordagem, quando aplicada ao estudo da gustação, o experimentador pretende caracterizar a relação entre topografias específicas de movimentos da face, incluindo lábios e língua, a concentrações e qualidades de substâncias químicas aplicadas à boca. Uma análise num nível ainda mais molecular na área de gustação envolve dissecar as características do comportamento de ingestão de líquidos, identificando separadamente aspectos como número de jorros de ingestão, duração dos jorros de ingestão, pausas entre jorros, volume ingerido, etc. Uma preocupação central nestas microanálises é a viabilidade de dissociar aspectos estritamente sensoriais de aspectos hedônicos da resposta ao estímulo químico (Crook, 1987).

Sugar de alta amplitude

Diferencia-se o uso de um procedimento baseado na resposta de sugar, originalmente proposto para o estudo da visão e posteriormente adaptado por Eimas e colaboradores (1971) para o estudo da audição. O que se segue é uma descrição baseada em Patricia Kuhl (1987) representativa do uso contemporâneo deste procedimento, chamado de sugar de alta amplitude. Dá-se à criança uma chupeta que contém, internamente, um transdutor de pressão. Desta forma monitora-se a ocorrência e a magnitude da resposta de sugar. Respostas de sugar de magnitude acima de um critério são reforçadas com a apresentação de um som. O som é mantido até que ocorra habituação da resposta, de acordo com um certo

critério de redução. Ocorrida a habituação, o som é mudado. A mudança no som gera desabituação da resposta. A desabituação é interpretada como indicador da discriminabilidade dos dois sons.

Kuhl aponta algumas limitações deste método: a) não separa convincentemente variáveis sensoriais de variáveis de atenção e motivação pelo estímulo; b) exige tratamento grupai de dados; c) crianças acima de 4 meses não demonstram interesse pela tarefa, o que limita seu uso para estudos longitudinais. Não obstante, parece ser o método disponível mais confiável e com resultados mais consistentes para o estudo da audição de crianças até os 4 meses.

Psicofisica da criança de 4 a 18 meses

Virar a cabeça reforçado

Este método, atribuído a Moore, Thompson e Thompson (1975), diferenciou-se a partir do comportamento reflexo de orientação da cabeça em direção à fonte de um som. A resposta de orientação é visualmente reforçada. Resumidamente, se a criança mantiver sua cabeça virada para afrente, algum tempo depois é apresentado um som; se a criança responder com uma virada de cabeça na direção do som, dentro de um critério de ângulo mínimo, ela é visualmente reforçada com a apresentação de imagens visuais projetadas (por exemplo, trechos de desenhos animados) ou a iluminação de um objeto interessante colocado nas imediações da fonte sonora. Este procedimento básico vem sofrendo refinamentos e adaptações para a investigação de aspectos específicos da audição da criança (vide, por exemplo, Kuhl, 1987; Sinnot, Aslin, 1985). Os aspectos interessantes deste método, segundo seus usuários, resumem-se no fato de que os dados são consistentes e por ser ele melhor que o método de sugar de alta amplitude, por permitir análise de dados individuais. Oferece também limitações: Kuhl relata dificuldade em manter crianças engajadas na tarefa a partir de 12 meses de idade.

Psicofisica da criança acima de dois anos

Procedimentos de condicionamento operante

Os procedimentos desenvolvidos para a psicofisica animal têm inspirado alguns autores no desenvolvimento de procedimentos que se prestem a crianças na faixa de 2 a 5 anos, idade a partir da qual procedimentos planejados para adultos começam a ser utilizáveis. Dobson (1990) examinou alguns procedimentos desenvolvidos dentro desta perspectiva para o estudo da visão. Os principais problemas específicos com os quais o pesquisador tem que lidar são a escolha de uma resposta e de um reforçador adequados à idade da criança e a escolha de um método psicofisico que atenda simultaneamente a critérios mais estritamente

psicofísicos de adequação ao problema em investigação e de estabilidade no desempenho da criança, o que às vezes pode significar uma densidade relativamente alta de reforço.

Equiparação intermodal

Este método foi originalmente desenvolvido para adultos, porém Stevens se preocupou em verificar a viabilidade de sua utilização com crianças e demonstrou que crianças de 5 anos, executando uma tarefa de equiparação entre intensidade luminosa e intensidade acústica, têm desempenho comparável ao de adultos na mesma tarefa. Esta demonstração é importante porque o procedimento é adequado para escalonamento, e os procedimentos para tal com adultos têm se valido do repertório de numerosidade para julgamento de magnitude sensorial, repertório este ainda não bem estabelecido nesta faixa de idade.

Avaliação da confiabilidade dos dados

A descrição acima deixa claro que há problemas na interpretação de dados psicofísicos obtidos da criança e, sobretudo, do infante. Banks e Dannemiller (1987) propõem quatro estratégias, as quais eles chamam de técnicas de verificação, para que se possa aumentar o grau de confiança de que estimativas de limiares reflitam o comportamento de mecanismos sensoriais e não de fatores não-sensoriais. Embora elas sejam apresentadas pelos autores no contexto específico de desenvolvimento visual, claramente podem ser adaptadas para o estudo de outros tipos de sensação e até mesmo para a pesquisa psicológica em outras áreas.

O uso de limiares como informação relativa

Banks e Dannemiller consideram a separação de variáveis sensoriais de não-sensoriais uma questão que tem que ser enfrentada se o problema em investigação envolver uma comparação absoluta entre crianças e adultos. A separação destas categorias de variável torna-se irrelevante se o problema em investigação envolver uma comparação em termos relativos entre crianças e adultos, como, por exemplo, a sensibilidade intra-idade para diferentes valores ao longo de alguma dimensão de estímulo. Neste segundo tipo de questão, variáveis não-sensoriais podem estar presentes — mas elas são uma constante — e pode-se obter funções com forma surpreendentemente equivalente para diferentes idades, a despeito de grandes diferenças nos valores absolutos.

Convergência de estímulo

Banks e Dannemiller definem convergência de estímulo como a habilidade de um modelo de mecanismo sensorial em predizer limiares obtidos para um tipo de estímulo a partir de limiares obtidos para outros tipos de estímulos, quando a mesma medida de resposta é usada com os dois tipos

de estímulo. Se as predições do modelo sensorial são acuradas, aumenta a confiança de que os dados reflitam variáveis sensoriais. Os autores consideram mais difícil interpretar a ausência de convergência de estímulo, já que ela pode ocorrer tanto em virtude da participação diferenciada de fatores não-sensoriais na interação com um dos estímulos como em virtude da escolha não apropriada de um modelo de mecanismo sensorial.

Convergência de resposta

Banks e Dannemiller definem convergência de resposta como a habilidade de se obter estimativas semelhantes de limiar usando medidas diferentes de resposta, quando os mesmos tipos de estímulo são usados para medidas diferentes de resposta. A observação de convergência de resposta permite afirmar que ela reflete os mesmos mecanismos sensoriais e, portanto, aumenta a confiabilidade dos dados de limiar. Esta estratégia é amplamente usada na comparação entre medidas comportamentais e eletrofisiológicas de limiar para as várias modalidades sensoriais.

Bom desempenho em relação a algum desempenho ideal conhecido

Banks e Dannemiller propõem que se compare o desempenho empírico da criança com o desempenho ideal teórico derivado a partir do conhecimento sobre o papel desempenhado por variáveis sensorialmente relevantes. Usando o exemplo dos próprios autores, seria possível comparar o desempenho empírico em uma tarefa de limiar para grades com o esperado a partir do que se sabe sobre as qualidades ópticas e a densidade de fotorreceptores. Embora os autores não explicitem, fica claro que a potência desta estratégia depende da disponibilidade de um bom modelo, para se poder interpretar adequadamente eventuais resultados negativos.

CONCLUSÕES

A psicofisica tem mantido os cientistas ocupados, embora as preocupações contemporâneas não sejam as de seu fundador. Ela ocupa um *status* científico diferenciado na Psicologia, por ser uma das poucas áreas nas quais há condições estabelecidas para quantificação. Do ponto de vista conceptual, é importante destacar a preocupação em relação a uma lei psicofisica geral e com poder preditivo. Do ponto de vista metodológico, observa-se um crescimento quanto às alternativas de tipos de observadores, humanos e não-humanos, a respeito dos quais pode-se obter dados fidedignos, graças ao contínuo desenvolvimento de estratégias de pesquisa. Fenômenos de sensação, percepção e cognição têm sido os principais objetos específicos de estudo da psicofisica. Nessas áreas observa-se uma trajetória no sentido de se estudar um número cada vez maior de categorias de fenômenos e fenômenos cada vez mais complexos.

REFERÊNCIAS BIBLIOGRÁFICAS

ADLER, H.E. Vicissitudes of Fechnerian psychophysics in America. In: RIEBER, R.W., SALZINGER, K. (Eds.). *Psychology, theoretical-historical perspectives.* New York: Academic Press, 1980. p.11-23.

BANKS, M.S., DANNEMILLER.J.L Infant visual psychophysics. In: SALAPATEK, P., COHEN, L (Eds.). *Handbook of infant perception.* From sensation to perception. Orlando (FL): Academic Press, 1987. v.1, p.115-184.

BERGAMASCO, N.H.P. Desenvolvimento de funções visuais básicas desde o nascimento. In: REUNIÃO ANUAL DA SOCIEDADE BRASILEIRA DE PSICOLOGIA. 22. *Resumos de comunicações científicas.* Ribeirão Preto: Sociedade Brasileira de Psicologia, 1992. p.397.

BIRNHOLTZ, J.C., BERNACERRAF, B.R. The development of human fetal hearing. *Science,* Washington. D.C., v.222, n.4623, p.516-518,1983.

BLOUGH, D.S. A method for obtaining psychophysical thresholds from the pigeon. *Journal of the Experimental Analysis of Behavior,* Bloomington (IN), v.1, p.31-43, 1958.

BLOUGH. D.S.. BLOUGH, P. Animal psychophysics. In: HONIG, W.K., STADDON, J.E. (Eds.). *Handbook of operant behavior.* Englewood Cliffs (NJ): Prentice-Hall, 1977. p.514-539.

BORING, E.G. *A history of experimental psychology.* 2.ed. Englewood Cliffs (NJ): Prentice-Hall. 1957.

COREN. S., PORAC. C, WARD. LM. *Sensation and perception.* New York: Academic Press, 1979.

COWART, B.J. Development of taste perception in humans: sensitivity and preference throughout the life span. *Psychological Bulletin,* Washington, DC. v.90. n.1. p.43-73, 1981.

CROOK, C. Taste and olfaction. In: SALAPATEK, P., COHEN. L. (Eds.). *Handbook of infant perception.* Orlando (FL): Academic Press, 1987. v.1: From sensation to perception, p.237-264.

DOBSON, V. Behavioral assessment of visual acuity in human infants. In: BERKLEY, M.A., STEBBINS. W.C. (Eds.). *Comparative perception.* New York: Wiley. 1990. v.1: Basic mechanisms, p.487-521.

DODWELL. P.C.. HUMPHREY, GK. MUIR, D.W. Shape and pattern perception. In: SALAPATEK, P., COHEN, L (Eds.). *Handbook of infant perception.* Orlando (FL): Academic Press, 1987. v.2: From perception to cognition, p. 1-77.

EIMAS, P.D.. SIQUELAND, E.R., JUSCZYK. P., VIGORITO, J. Speech perception in infants. *Science,* Washington, D.C., v.171, p.303-306,1971.

FALMAGNE, J. C. *Elements of psychophysical theory.* New York: Oxford University Press, 1985.

GESCHEIDER, G.A. *Psychophysics:* method and theory. New York: Wiley, 1976.

------------- . Psychophysical scaling. *Annual Review of Psychology,* v.39, p.169-200, 1988.

GREEN, D.M., SWETS, J.A. *Signal detection theory and psychophysics.* New York: Wiley, 1966.

KRUEGER. L.E. Reconciling Fechner and Stevens: towards a unified psychophysical law. *Behavioral and Brain Sciences,* Cambridge, v.12. p.251-320. 1989.

KUHL, P.K. Perception of speech and sound in early infancy. In: SALAPATEK, P., COHEN, L. (Eds.). *Handbook of infant perception.* Orlando (FL): Academic Press, 1987. v.2: From perception to cognition, p.275-382.

LUCE, R.D. "On the possible psychophysical laws" revisited: remarks on cross-modal matching. *Psychological Review,* Washington, D.C., v.97, p.66-77, 1990.

LUCE, R.D., KRUMHANSL.C.L Measurement scaling, and psychophysics. In: ATKINSON, R., HERRNSTEIN, R., LINDZEY, G., LUCE, R.D. (Eds.). *Stevens'handbook of experimental psychology.* 2.ed. New York: Wiley, 1987. v.1: Perception and motivation, p.3-74.

MARSHALL, L, BRANDT, J.F. The relation between loudness and reaction time in normal hearing listeners. *Acta Oto-Laryngologica,* Stockholm, v.90, p.244-249, 1980.

MOORE, J.M., THOMPSON, G., THOMPSON, M. Auditory localization of infants as a function of reinforcement conditions. *Journal of Speech and Hearing Disorders,* Rockville (MD). v.40, p.29-34, 1975.

PFINGST, B.E., HIENZ, R., KIMM, J., MILLER, J. Reaction time procedure for measurement of hearing: suprathreshold functions. *Journal of the Acoustical Society of America,* v.57, p.421-431, 1975.

REASON, J.T. Individual differences in auditory reaction time and loudness estimation. *Perceptual and Motor Skills,* Missoula (MT), v.26, p. 1089-1090. 1968.

REISMAN, J.E. Touch, motion, and proprioception. In: SALAPATEK, P., COHEN, L (Eds.). *Handbook of infant perception.* Orlando (FL): Academic Press, 1987. v. 1 : From sensation to perception, p.265-303.

SCHNEIDER, B.A., TREHUB, S.E., BULL, D. The development of basic auditory processes in infants. *Canadian Journal of Psychology,* Ottawa, v.33, p.306-319, 1979.

SINNOT, J.M., ASLIN, R.N. Frequency and intensity discrimination in human infants and adults. *Journal of the Acoustical Society of America,* v.78, n.6, p.1986-1992,1985.

STEBBINS, W.C. (Ed.). *Animal psychophysics.* The design and conduct of sensory experiments. New York: Appleton-Century-Crofts, 1970.

STEBBINS, W.C. Perception in animal behavior. In: BERKLEY. M.A., STEBBINS, W.C. (Eds.). *Comparative perception.* New York: Wiley, 1990. v.1 : Basic mechanisms.

STEBBINS, W.C, MILLER, J. Reaction time as a function of stimulus intensity for the monkey. *Journal of the Experimental Analysis of Behavior,* Bloomington (IN), v.7, p.309-312,1964.

STEVENS, S.S. *Psychophysics:* introduction to its perceptual, neural and social prospects. New York: Wiley, 1975.

THOMPSON, R.F., ROBINSON, D.N. Physiological psychology. In: HEARST, E. (Ed.). *The first century of experimental psychology.* Hillsdale (NJ): Lawrence Erlbaum, 1979. p.407-454.

CAPÍTULO 3

MEDIDA PSICOMETRIA

Luiz Pasquali
Instituto de Psicologia
Universidade de Brasília

Há cerca de 20 anos, Guttman (1971) ainda se interrogava o que exatamente significava "medida" em ciências sociais e do comportamento. Embora, nestas ciências, aparecessem correntes os termos sociometría, antropometria, biometria, psicometria, econometria e outras 'metrias', continuavam dúvidas sobre sua significação no campo da epistemologia e da metodologia. Os vários prefixos das "...metrias" evidentemente revelavam a área de conteúdo em que a medida era aplicada. Assim, psicometria seria o uso da medida em Psicologia.

Esta situação levantada por Guttman não está de todo resolvida ainda hoje. De fato, a teoria da medida em ciências não constitui campo pacífico entre os pesquisadores, sobretudo em ciências do comportamento. Outro complicador, neste contexto, é a tendência de alguns em reduzir, por exemplo, psicometria, cuja preocupação central é a construção e verificação de hipóteses científicas, à psicoestatística, cuja preocupação é a inferência a partir de amostras. Aliás, este tipo de divergência foi o que provocou, em análise fatorial, a divisão do grupo de Thurstone dos anos 30 em várias correntes, cada qual seguindo seus interesses pessoais de psicometristas, de estatísticos ou de matemáticos, inclusive com a criação de revistas especializadas divergentes da *Psychometrika.*

Este capítulo pretende caracterizar a psicometria dentro de uma orientação epistemológica quantitativista, mas como ramo das ciências empíricas e não das matemáticas. Estas duas não são conflituosas, mas são epistemológicamente independentes. A distinção precisa ser defendida — e que seja também cobrado o ônus da prova para a justificativa da viabilidade da associação das duas, isto é, Ciência, de um lado, e Matemática, do outro (vide cap. 1). Este capítulo se divide em duas partes: a teoria psicométrica e os parâmetros psicométricos da medida.

Em seu sentido etimològico, psicometria seria, conforme insinuou Guttman (1971), toda a classe de medida em Psicologia, similarmente a sociometría ser na Sociologia, econometria na Economia, etc. Em seu sentido mais restrito — e é neste que ela é normalmente entendida —, psicometria constitui uma das várias formas de medição em Psicologia. Ela é uma das formas de medida por teoria (vide cap. 1), onde se situam igualmente a teoria dos jogos e da detecção do sinal.

A teoria que fundamenta a psicometria neste sentido estrito, além de assumir os postulados da teoria da medida em geral, trabalha com o modelo da estrutura latente (traços psicológicos, *latent modeling).* Neste contexto, alguns conceitos-chave precisam ser explicitados: traço latente, sistema, propriedade, magnitude, bem como a representação comportamental da estrutura latente.

Traço latente

O conceito de traço latente não é isento de ambigüidades e controvérsias entre os autores que trabalham com tal construto. A variedade de expressões utilizadas para representá-lo já indica tal dificuldade. Traço latente vem referido sob denominações como: variável hipotética, variável fonte, fator, construto, conceito, estrutura psíquica, traço cognitivo, processo cognitivo, processo mental, estrutura mental, habilidade, aptidão, traço de personalidade, processo elementar de informação, componente cognitivo, tendência e outros. A própria natureza ontológica de traço latente deixa dúvidas se ele é concebido como um rótulo, representando uma síntese hipotética de um conjunto de comportamentos reais, ou como uma realidade mental. Para este autor, o conceito faz mais sentido quando entendido como realidade na concepção popperiana de que é real aquilo que age sobre coisas consideradas reais, como as coisas físicas materiais: "Deve-se então admitir que as entidades reais podem ser concretas ou abstratas em vários graus. Em Física, aceitamos forças e campos de força como reais, pois agem sobre coisas materiais. Mas essas entidades são mais abstratas e, talvez, também mais conjeturáis ou hipotéticas do que são as coisas materiais comuns. Forças e campos de força são ligados a coisas materiais, a átomos e a partículas. Têm um caráter dispositivo: são tendências para interagir. Podem assim ser descritas como entidades teóricas altamente abstratas, nós as aceitamos como reais, quer elas ajam de forma direta ou indireta sobre as coisas materiais." (Popper, Eccles, 1977, p.27-28).

Além desta controvérsia, existem diferentes maneiras de conceber traço latente quando se trata de definir sua estrutura elementar. Na verdade, há aqui duas tendências distintas e em vários níveis: concepção elementarista (reducionista) e concepção estrutural, conforme detalhado na Figura 3-1.

Fig. 3-1 — Visões elementarista e estrutural de traço latente.

Há os que entendem traço latente como uma estrutura global, seja constituindo toda a psique do ser humano ou conjuntos dela. No caso das aptidões humanas, por exemplo, Spearman (1904) defende a teoria do fator **intelectual** único (fator G); Piaget (1952) fala do desenvolvimento das **estruturas** cognitivas. Na área da emoção e da motivação aparecem as **tipologias de tipo** Jung (1921), Kretschmer (1925) e Sheldon (1940, 1942). **Estas** concepções apresentam uma tendência de considerar os traços **latentes** como grandes estruturas que variam de sujeito para sujeito. **Dentro** ainda de uma concepção estrutural, outros autores concebem os **traços** latentes de uma forma mais diferenciada, quando falam de fatores. É **a** tradição na orientação da análise fatorial em psicometria, onde os **fatores** são concebidos como variáveis-fonte responsáveis pela qualidade da execução das tarefas comportamentais. Embora pareçam já elementares, os fatores apresentam ainda um caráter globalizante, dado que não expressam processos cognitivos elementares, mas sim um possível conjunto destes que são necessários para a execução de uma tarefa concreta. Pelo menos, esta é a crítica que Sternberg (1977) faz desta concepção fatorista do traço latente. Sternberg, na verdade, concebe traço latente como algo elementar, isto é, o último elemento cognitivo a que se pode reduzir uma atividade cognitiva, os processos cognitivos (o autor trabalha na área das aptidões). A teoria do processamento da informação (Newell, Simon, 1972) leva ainda mais longe este elementarismo, defendendo o conceito de *elementary information process* (eip) como sendo o processo mais elementar possível no processamento da informação, o qual não pode ser analisado em elementos menores.

Para ilustrar, talvez de forma grosseira, estas várias concepções, no caso dos processos cognitivos ou das habilidades e aptidões, podemos considerar a Tabela 3-1, na página seguinte.

Assim, as concepções de traço latente dependem do nível de especificidade que se quer dar a este construto ou parâmetro. Os fatoristas estão mais interessados em chamar de traço latente aquele conjunto de processos cognitivos necessários para a execução de uma tarefa (de fato,

Tabela 3-1 — Enfoques conceituais de processo cognitivo

ENFOQUE	REFERÊNCIA	TRAÇO LATENTE	ILUSTRATIVA
Processamento da informação	elementary information process — eip	elemento atômico da Física Nuclear	Newel & Simon, 1972
Psicologia Cognitiva	componente cognitivo	elemento da tabela periódica de Química	Sternberg. 1977
Psicometria	fator	elemento natural (Geologia. Geografia.	Fatoristas (Thurstone Cattell. Guilford ...)

um barramento correlacionado de processos), falando de habilidades primárias, que seriam combinações de processos cognitivos elementares, isto é, de representações mentais de objetos e símbolos. O fator seria um sistema de processos cognitivos ou de componentes cognitivos. Ao contrário, Sternberg chama de processo cognitivo estas mesmas representações mentais individuais, que serão os componentes cognitivos. Agora, para representar mentalmente objetos e símbolos, uma série de processos mais elementares ainda são necessários e, então, estes sim, seriam finalmente os processos elementares básicos do processamento da informação para Newell e Simon, elementos que se combinam num sistema de processos da informação para a explicação de uma tarefa comportamental. Onde parar nesta tendência reducionista? Sternberg (1977, p.65-66) afirma: "o componente não é necessariamente, e normalmente não o é, a unidade mais elementar de comportamento que se possa estudar. Operações que são consideradas sem importância dentro da teoria são especificadas no modelo do processamento da informação do desempenho de uma tarefa, mas não serão identificadas como componentes separados. A razão para esta seletividade é que tarefas complexas podem requerer centenas ou até milhares de operações, a maioria das quais se apresentam desinteressantes do ponto de vista da teoria".

Além da diferença no nível de reducionismo, outra vertente importante de diferenças entre estes vários sistemas de conceber o traço latente consiste na visão mais estruturalista das concepções holísticas, que tendem a considerar os traços latentes como entidades; as concepções mais elementaristas tendem a considerar traço latente como processos. Assim, Newell e Simon consideram como "eips" a discriminação, a testagem e a comparação, a criação de símbolos, a criação de estruturas de símbolos, produção de respostas externas em função de estruturas simbólicas internas, designação de estruturas simbólicas e memorização de estruturas simbólicas. Por sua vez, Sternberg fala de processos de codificar, inferir, mapear, aplicar, justificar e responder. As diferenças

individuais que ocorrem nestes processos seriam devidas à dificuldade e duração que diferentes sujeitos encontram ou necessitam para eliciar estes processos, enquanto para os fatoristas, por exemplo, as diferenças surgiriam principalmente em função da magnitude (tamanho, dimensão, quantidade) do traço latente possuído por diferentes sujeitos. A psicometria trabalha com o conceito fatorista de traço latente.

Sistema

O sistema representa o objeto de interesse, chamado também de objeto psicológico. A psicometria enfoca como seu objeto específico as estruturas latentes, os traços psicológicos; ela teoriza a partir destas estruturas hipotéticas. Deste enfoque, evidentemente, surgem dificuldades, dado que a ciência empírica, dentro da qual a Psicologia se define, tem como objeto de conhecimento os fenômenos naturais abordados através da observação, que, no caso da Psicologia, é o comportamento. Este problema será abordado na seção da representação comportamental da estrutura latente. Aqui é relevante salientar que a psicometria trabalha com a teoria dos traços latentes, sendo, portanto, as estruturas psicológicas latentes o seu objeto ou sistema direto de interesse. O sistema pode ser considerado de vários níveis, dependendo do interesse do pesquisador. Poder-se-ia falar de um sistema universal e de sistemas locais. O universal sendo a estrutura psicológica total do ser humano, e os sistemas locais, os vários subsistemas de interesse. Assim, a inteligência pode ser considerada um subsistema dos processos cognitivos, e estes, da estrutura latente geral; ou mesmo, a inteligência, digamos, verbal, pode ser considerada um sistema quando ela for o interesse imediato e na qual vários aspectos podem ser considerados, como a compreensão verbal e a fluencia verbal. Sistema, portanto, constitui-se como tal quando representa o objeto imediato de interesse dentro de um delineamento de estudo e não é uma entidade ontológica, monolítica e unívoca.

Propriedade

Um sistema apresenta atributos que são os vários aspectos ou as propriedades que o caracterizam. Por exemplo, o sistema físico se apresenta com os atributos de massa, comprimento, etc. Similarmente, a psicometria concebe os seus sistemas como possuidores de propriedades/atributos que definem os mesmos, sendo estes atributos o foco imediato de observação/medida. Assim, a estrutura psicológica apresenta atributos do tipo processos cognitivos, processos emotivos, processos motores, etc. A inteligência, como subsistema, pode apresentar atributos de tipo raciocínio verbal, raciocínio numérico, etc. O sistema se constitui como objeto hipotético que é abordado (conhecido) através da pesquisa de seus atributos.

Magnitude

A psicometria assume, ainda, que estes atributos psicológicos apresentam magnitude: os atributos são dimensões, isto é, são mensuráveis. Trata-se do conceito de quantidade: os atributos ocorrem com quantidades definidas e diferentes de indivíduo para indivíduo. Quantidade é um conceito matemático que se define em função dos axiomas de ordem e de aditividade dos números: os números não somente são diferentes, mas uns são maiores que outros, de sorte que eles podem ser ordenados numa série monotônica crescente de magnitude. Ao se falar de magnitude dos atributos empíricos, quer se referir, pelo menos, a esta propriedade numérica de ordem crescente. Digo pelo menos, porque nem sempre é possível salvar na medida o axioma da aditividade que implica a possibilidade de concatenação, resultando em medida de nível intervalar ou de razão. Aliás, é esta suposição de magnitude das propriedades psicológicas que torna interessante a utilização do modelo matemático no estudo dos fenômenos de que trata a Psicologia.

O problema da representação comportamental

Mesmo se admitindo que as estruturas latentes tenham atributos e que estes possuam magnitude, fica o problema fundamental de que estes atributos são impervios à observação empírica, que é o método da Ciência. Então, como fica a utilidade de todo este teorizar? Estamos aqui nos defrontando com o problema da representação: qual é a maneira adequada de se representar estes atributos latentes para que possam ser cientificamente abordados? Embora o problema pareça, e é na verdade, grave, ele não é específico da psicometria; eie ocorre na própria Física, com a teoria quàntica, por exemplo.

Como o comportamento (verbal, motor) é o único nível em que se pode trabalhar cientificamente (empiricamente) em Psicologia, é neste nível que se deve procurar a solução para o problema da representação e, portanto, do conhecimento dos processos latentes. Está ali também o problema básico da psicometria — a legitimidade de suas operações depende da legitimidade desta representação. A teoria que fundamenta o isomorfismo comportamento-processos latentes é o fulcro epistemológico da psicometria, juntamente com a concepção de processos latentes como dimensões, isto é, atributos mensuráveis. Postula-se que, ao se operar sobre o sistema comportamento, está-se operando sobre os traços latentes (isomorficamente). Assim, a medida que se faz em nível comportamental é a medida dos traços latentes.

Como o comportamento representa estes traços latentes? É o problema das definições operacionais. A psicometria responde a esta questão pela análise de uma série de parâmetros que os comportamentos

(tipicamente chamados de itens) devem apresentar. Tais parâmetros são mencionados a seguir:

Modalidade

Em termos de seu conteúdo, os comportamentos (itens) podem ser de tipo verbal ou motor. Dentro destes, podem-se distinguir outros. No caso do verbal, por exemplo, o item pode ser verbal propriamente, numérico, espacial, abstrato, etc, dependendo do conteúdo semântico sobre o qual o comportamento opera ser de palavras, números, dimensões espaciais, etc. Pode ser também mais ou menos abstrato, dependendo do nível de universalidade dos conceitos envolvidos: conceitos singulares, universais de menor abstração, universais de maior abstração. Neste particular, a psicometria deveria interagir com a psicolinguistica, já que apresenta interface com a questão do significado.

Saturação

O comportamento humano tipicamente se apresenta como multimotivado, dado que fatores múltiplos entram na sua aparição, sendo, portanto, difícil, senão impossível, determinar causas ou fatores únicos para qualquer comportamento, ao menos de adultos. Isto implica que seria impossível definir comportamentos (itens) críticos para qualquer traço latente, no sentido de um comportamento 'x' ser específico e único de tal traço e não tendo interface com qualquer outro traço. Podemos dizer, então, que somente parte do comportamento 'x' representa o traço, ele covaria com o traço; mas esta covariação não constitui toda a variância do 'x'. É, por isso, importante descobrir o nível desta covariancia 'x' e traço latente em questão. Tipicamente tal covariancia se expressa estatisticamente através da sua carga ou saturação fatorial, que pode variar de zero a um (positivo ou negativo), sendo que, no caso de ser zero, o comportamento seria uma representação equivocada, inadequada, do traço. Este parâmetro se relaciona à questão da unidimensionalidade das escalas de medida.

Dificuldade (complexidade)

Um comportamento é mais dificil ou mais complexo à medida que ele exige maior nível de magnitude do traço em questão para ser eficaz ou corretamente executado. A expressão 'dificuldade' se originou dentro da medida das aptidões e é mantida, por exemplo, no parâmetro 'b' da *Item Response Theory* (IRT), mesmo quando se trata da medida de atitudes ou traços de personalidade em geral. Talvez a denominação 'complexidade' fosse mais adequada para representar este parâmetro, de uma vez que ela especifica que um comportamento é mais complexo e, portanto, mais difícil, porque a sua correta execução (no caso de se tratar de aptidão cognitiva) ou a adesão a seu conteúdo semântico (no caso de traços de

personalidade e atitudes) depende de um maior nível de magnitude no traço latente. O que exatamente torna um item mais complexo é ainda tema de pesquisa, do qual a Psicologia Cognitiva vem se interessando bastante como forma de estudar os processos cognitivos. A psicometria avalia este parâmetro através de técnicas puramente estatísticas, mas seria de enorme valor a descoberta dos elementos que constituem maior complexidade no item, sobretudo para fins de construção do próprio elenco de itens da medida dos traços latentes. Este parâmetro afeta a questão da amplitude de uma escala de medida: o elenco de itens cobre adequadamente toda a extensão de magnitudes possíveis de um dado traço ou somente um segmento delas e qual segmento?

Discriminação

O poder discriminativo de um item (comportamento) se define como a capacidade que ele apresenta de separar (discriminar) sujeitos com magnitudes próximas do mesmo traço. Quanto mais extremas devam ser as magnitudes do atributo para que o item possa discriminá-las, menos discriminativo ele é e vice-versa. A IRT define como 'a' este parâmetro. Que característica do item determinaria seu poder discriminativo? Novamente a Psicologia Cognitiva poderia lançar luzes nesta questão, definindo os elementos cognitivos que a reação **a** um item utiliza. Seria a univocidade semântica do item, isto é, um sentido bem definido com nível reduzido de ruído, a saber, conceitos despojados de conotações? Uma informação desta natureza auxiliaria grandemente a construção de itens comportamentais mais típicos e adequados para a medida dos traços.

Viés de resposta

Mesmo apresentando bons índices nos parâmetros acima descritos, há toda uma série de dificuldades que aparecem afetando a qualidade da resposta do sujeito aos itens, dificuldades estas que provêm de fatores subjetivos do respondente e que poderiam ser agrupadas dentro do conceito de tendências. Tendência seria uma atitude, consciente ou não, **de o** sujeito responder de maneiras sistemáticas alheias ao conteúdo semântico dos itens — erros de resposta ao responder ao acaso, respostas estereotipadas (sempre nos extremos de uma escala ou **no** ponto neutro), respostas em função de supostas expectativas dos outros (desejabilidade social) ou em função de uma idéia preconcebida sobre **o** objeto de avaliação (efeito **de** halo), etc. Vários destes problemas podem **ser** parcialmente evitados, **se se** puder desvendar os fatores sistemáticos responsáveis pelas respostas estereotipadas. Assim, **a IRT** é **capaz de contornar o problema das respostas** dadas ao acaso **(parâmetro** 'c'); **o formato das escalas de resposta pode** reduzir **a** ocorrência **de erros do tipo respostas extremadas ou neutras, etc.**

O parâmetro fundamental da medida psicométrica (escalas, testes...) é a demonstração da conformidade da representação, isto é, a demonstração do isomorfismo entre a ordenação nos procedimentos empíricos e a ordenação nos procedimentos teóricos do traço latente. Significa demonstrar que a operacionalização do atributo latente em comportamentos (¡tens) de fato corresponde a este atributo. Esta demonstração é tipicamente tentada através de análises estatísticas dos ¡tens individualmente e da escala em seu todo. Infelizmente a literatura, neste particular, não mostra muita preocupação com a formulação de uma teoria clara, muito menos axiomatizada, sobre o atributo que permitisse uma elaboração mais bem delineada e planejada de uma escala de comportamentos pertinentes ao atributo. Possivelmente esta situação se deve 1) à predominância de um enfoque positivista baseado quase exclusivamente na análise de um elenco de ¡tens, coletado mais ou menos ao acaso ou intuitivamente, em vez de uma pesquisa dos elementos cognitivos envolvidos nos processos do atributo psicológico e, também, 2) ao fato de que o desenvolvimento da psicometria tem sido preponderantemente viabilizado por pesquisadores cuja formação e preocupações eram mais de estatísticos do que de psicólogos. O desenvolvimento da pesquisa da Psicologia Cognitiva, particularmente do tipo Sternberg (1977, 1979,1980) e das pesquisas feitas no centro de Pittsburgh (Mulholland, Pellegrino, Glaser, 1980; Pellegrino, Mumaw, Shute, 1985; Carpenter, Just, Shell, 1990), deverá auxiliar substancialmente para remediar ou resolver este problema. Os trabalhos de Guilford (1959) também devem ser mencionados neste particular. No momento, em psicometria, se insiste ainda de maneira exclusiva numa solução estatística. Por outro lado, as contribuições que a Psicologia Cognitiva tem, no momento, a dar nesta área da instrumentação psicométrica são ainda muito precárias para servir de base na elaboração e análise dos instrumentos psicológicos.

De qualquer forma, a comunidade científica desenvolveu uma série de parâmetros mínimos que a medida psicométrica deve apresentar para se constituir instrumento legítimo e válido. Os parâmetros mais básicos se referem à análise dos itens (dificuldade e discriminação) e à validade e confiabilidade do instrumento.

Análise dos itens

Há dois tipos de análise de itens, que poderíamos chamar de análise teórica e análise empírica ou estatística.

Análise teórica dos itens

Esta análise é feita por juizes e visa estabelecer a compreensão dos itens (análise semântica) e a pertinência dos mesmos ao atributo que

pretendem medir. Esta última é, às vêzes, chamada de análise de conteúdo, mas propriamente deve ser chamada de análise de construto, dado que precisamente procura verificar a adequação da representação comportamental do(s) atributo(s) latente(s).

No caso da análise semântica, duas preocupações são relevantes: primeiro, verificar se os itens são inteligíveis para o estrato mais baixo (de habilidade) da população-meta e, por isso, a amostra para esta análise deve ser feita com este estrato; segundo, para evitar deselegância na formulação dos itens, a análise semântica deverá ser feita também com uma amostra mais sofisticada (de maior habilidade) da população-meta (para garantir a chamada Validade aparente' do teste). De qualquer forma, a dificuldade na compreensão dos itens não deve se constituir fator complicador na resposta dos indivíduos, dado que não se quer medir a compreensão deles (a não ser, obviamente, que o teste queira medir precisamente isto), mas sim a magnitude do atributo a que os itens se referem.

Na análise de conteúdo, os juizes devem ser peritos na área do construto, pois sua tarefa consiste em ajuizar se os itens estão se referindo ou não ao traço em questão. Uma tabela de dupla entrada, com os itens arrolados na margem esquerda e os traços no cabeçalho, serve para coletar esta informação. Uma concordância de, pelo menos, X0% entre os juizes pode servir de critério de decisão sobre a pertinência do item ao traço a que teoricamente se refere.

Análise empírica dos itens

A análise da dificuldade e da discriminação dos itens se faz em cima dos dados coletados de uma amostra de sujeitos, utilizando-se de análises estatísticas.

1) Dificuldade dos itens

A dificuldade do item é definida em termos da porcentagem (proporção) de sujeitos que dão respostas corretas (testes de aptidão) ou de acordo/preferência (testes de personalidade) ao item. Assim, sobre um item que é respondido corretamente ou aceito por 70% dos sujeitos é afirmado ser ele mais fácil que um outro que recebeu 30% de respostas corretas.

Qual a dificuldade ideal dos itens de uma escala ou teste? A resposta depende da finalidade do teste. Se se deseja um teste para selecionar os melhores ou para determinar se um patamar 'x' de conhecimento foi atingido (como nos testes de referência a critério), então os itens devem todos apresentar o nível de dificuldade do patamar que se quer como critério de seleção. Assim, se se deseja selecionar somente os 30% melhores candidatos, os índices de dificuldade dos itens devem ser em

torno de 30%. Se, entretanto, o interesse consiste em avaliar a magnitude diferencial dos traços nos sujeitos, como geralmente é o caso, então uma distribuição mais equilibrada dos itens em termos de dificuldade é requerida. Neste caso, o interesse se centra sobre o poder de um teste discriminar diferentes níveis de habilidades, atitudes, etc, nos sujeitos e, por conseguinte, os itens devem poder avaliar tanto os que possuem pouca quanto muita habilidade. Entretanto, é bom saber que itens que todos os sujeitos acertam ou igualmente aceitam e itens que ninguém acerta ou aceita são ¡tens inúteis para fins de diferenciar indivíduos; de fato, tais itens não trazem nenhuma informação. Os itens que trazem maior informação são aqueles cujo índice de dificuldade se situa em torno de §0%, pois, neste caso, 50% dos sujeitos acertam e 50% erram, resultando 50 x 50 = 2.5(X) comparações possíveis, ao passo que um item com dificuldade 30% teria 70% de erros e 30% de acertos, resultando num nível de 30 x 70 = 2.100 bits de informação. Obviamente, um item com dificuldade 100% ou 0% produzirá zero informação. Deve-se concluir daí que todos os itens de um teste devam ter dificuldade 50%? Embora a maioria dos itens devam apresentar tal índice de dificuldade, nem todos o deverão, pois que assim poder-se-ia discriminar apenas dois niveis da magnitude do traço medido, dado que itens com o mesmo nível de dificuldade terão altas intercorrelações, determinadas pela circunstância de que serão os mesmos sujeitos que sempre acertam ou sempre erram os itens todos. Haveria, então, uma distribuição mais adequada dos itens de um teste em termos de dificuldade? Considerando que eles devem cobrir toda a extensão de magnitude do traço e que os itens de dificuldade 50% são os que produzem maior informação, pode-se sugerir que uma distribuição dos mesmos mais ou menos dentro de uma curva normal seria o ideal. Assim, se considerarmos a amplitude de um atributo ou traço numa escala de 100 pontos, podemos dividi-la em cinco níveis de magnitudes: 0 a 20, 20 a 40, 40 a 60, 60 a80 e 80 a 100, distribuindo os itens assim: 10% deles em cada uma das duas faixas extremas, 20% em cada uma das duas faixas seguintes e 40% na faixa média (vide Figura 3-2).

% DE ACERTOS	FAIXA	NÚMERO DE ITENS
80 a 100	V	10%
60 a 80	IV	20%
40 a 60	III	40%
20 a 40	II	20%
0 a 20	I	10%

Fig. 3-2 — Distribuição ideal dos itens por índice de dificuldade.

2) Discriminação dos itens

Discriminação se refere ao poder de um item em diferenciar sujeitos com magnitudes diferentes de traço do qual o item constitui a representação comportamental. Quanto mais próximas forem as magnitudes do traço que o item puder diferenciar, mais discriminativo eie é. Assim, Poder-se-ia dizer que discriminação se refere ao poder que o item possui de diferenciar sujeitos com magnitudes próximas do traço a que se refere.

a) Grupos-critério

A dificuldade envolvida na tarefa de avaliar o poder discriminativo dos itens consiste na escolha dos sujeitos que servirão de base como grupos-critério que o item deve diferenciar. A escolha dos critérios para efetuar a análise da discriminação dos itens tem dependido, na prática, dos objetivos do teste. Assim existem critérios externos e critérios internos ao próprio teste cujos itens se quer analisar. Critérios externos para estabelecer os grupos-critério podem ser, por exemplo, sujeitos psiquiátricos e sujeitos não-psiquiátricos para avaliar o poder de discriminação dos itens em testes psiquiátricos, ou sujeitos que tiveram êxito e sujeitos que fracassaram num curso de treinamento, ou, ainda, tipos de ocupações, etc. Enfim, trata-se de estabelecer grupos que se diferenciam em algum comportamento definido como relevante com referência aos objetivos do teste e verificar se os itens do teste são capazes de, individualmente, diferenciá-los.

Utilizam-se também critérios internos ao próprio teste para definir estes grupos-critério. Tipicamente é escolhido o escore total no próprio teste para determinar os grupos extremos de sujeitos: grupo superior e grupo inferior. Em amostras grandes, selecionam-se os 27% superiores e os 27% inferiores para comporem os dois grupos (Kelley, 1939). Evidentemente, em amostras menores, este percentual deverá ser maior, visto que os grupos de comparação devem apresentar um número suficiente de sujeitos para permitir análises estatísticas válidas. De modo geral, algo em torno de 30% será adequado; contudo, em amostras normais e grandes é costumeiro se utilizar a "regra 27%", como ficou sendo conhecida.

b) índices de discriminação

Existem dezenas de técnicas estatísticas para estabelecer o índice de discriminação do item (Anastasi, 1988), os quais produzem basicamente resultados similares (Oosterhof, 1976). Os mais utilizados são a análise da diferença de médias ou de percentagens dos sujeitos que passaram (testes de aptidão, onde há respostas certas e erradas) ou aceitaram (testes de personalidade, atitude) o item no grupo superior *vis-à-vis* o grupo inferior, bem como coeficientes de correlação, especialmente o coeficiente *phi(ty)* e a correlação bisserial.

• O índice D

É um dos mais fáceis para ser computado, porque consiste simplesmente na diferença de porcentagens de acertos no grupo superior e no grupo inferior, isto é, S - I ou, em inglês, U - L (ULI ou ULD). Veja exemplo na Tabela 3-2.

Tabela 3-2 — Computação do índice D

Item	% dos que passaram		Ìndice D
	Grupo superior	Grupo inferior	
1	80	40	40
2	100	90	10
3	30	50	-20
4	55	55	0
5	75	40	35

O índice D tem que ser positivo, e quanto maior for, mais discriminativo será o item. Obviamente, um D nulo ou negativo demonstra ser o item não-discriminativo.

• O teste "t"

Um índice de discriminação mais exato, embora mais laborioso de se conseguir, consiste na análise da diferença entre as médias obtidas pelos grupos superior e inferior. Neste caso, é necessário o cálculo das respectivas médias e de suas variâncias. O nível de significancia do teste 't' pode ser verificado com exatidão em tabelas estatísticas próprias.

• Coeficientes de correlação

O coeficiente *phi* trabalha com dados dicotômicos, numa tabela de quatro cáselas, e produz um valor que vai de -1 a +1, como qualquer coeficiente de correlação. O cálculo deste coeficiente pode ser conseguido através das tabelas de Jurgensen (1947), para o caso em que o número de sujeitos for o mesmo nos dois grupos de sujeitos, ou das tabelas de Edgerton (I960), para o caso deste número ser diferente.

O coeficiente bisserial de correlação (rb) utiliza as médias dos escores dos sujeitos que passaram ou que não passaram o item. Os livros de estatística apresentam várias formas de cálculo deste coeficiente (Guilford, Fruchter, 1973).

O cálculo do índice de discriminação com base no escore total do teste apresenta um problema teórico. Na verdade, procura-se analisar a adequação do item (em termos de discriminação) baseada nas infor-

mações obtidas de todo o elenco de ¡tens (escore total). Tal procedimento parece incongruente, dado que a adequação dos demais itens também está por ser demonstrada, inclusive a esta altura das análises do teste ainda não se sabe se os itens do teste são homogêneos, isto é, se o teste é unidimensional, suposição necessária para se poder obter um escore total. Tenta-se resolver este problema procedendo-se a uma análise fatorial dos itens antes da própria análise individual dos mesmos.

3) Item Response Theory (IRT ou ICC)

Esta teoria de análise dos itens possui uma longa história (desde os anos 50), mas só bem recentemente está sendo utilizada praticamente, dadas as dificuldades das análises estatísticas envolvidas que exigem o poder de avançados *softwares* e micros de certo porte. AIRT é uma teoria estatística, mas de utilização direta na psicometria, inclusive com grandes vantagens sobre outros métodos tradicionais de análise de itens.

Os modelos matemáticos envolvidos nesta teoria são bastante complexos e, embora seu conhecimento constitua uma grande vantagem, não parece ser indispensável seu domínio para uso inteligente por parte do usuário. A compreensão da sua lógica, entretanto, é indispensável.

Primeiramente, há vários modelos matemáticos envolvidos na IRT. Na verdade, há três deles principais, dependendo do número de parâmetros que pretendem avaliar dos itens. Os parâmetros em questão são a dificuldade, a discriminação e a resposta aleatória (ou melhor, a resposta correta dada ao acaso). Assim, temos o modelo logístico de 1, 2 ou 3 parâmetros.

Todos os modelos trabalham com traços latentes, isto é, teorizam sobre as estruturas latentes, como faz a teoria psicométrica. Entendem os sistemas psicológicos latentes como possuindo dimensões, isto é, propriedades de diferentes magnitudes ou mensuráveis. Por isso, esta teoria também é conhecida como a teoria do traço latente ou a teoria da curva característica do item *{¡tem characteristic curve* - ICC). A teoria supõe que o sujeito possui um certo nível de magnitude do traço latente, designado por *theta* (0), o qual é determinado através da análise das respostas dos sujeitos por meio de diversas funções matemáticas. A função do modelo completo de três parâmetros é:

$$P_i(\theta) = c_i + (1 - c_i)\frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}}$$

A probabilidade de resposta correta, que define a posição (0) do indivíduo no traço medido, é função de três parâmetros: a corresponde ao índice de discriminação do item e é determinado pela curvatura da ogiva no ponto de inflexão; b é o parâmetro da dificuldade/preferência e é

expresso pelo valor no eixo dos X no ponto de inflexão da curva; e é o parâmetro que determina as respostas acertadas/preferidas por acaso, sendo o D uma constante usualmente com valor 1,7.

Os três modelos de IRT mais conhecidos são os seguintes: 1) o modelo logístico de um parâmetro ou o modelo Rasch (1966) — Rasch faz a suposição de que os itens possuem o mesmo nível de discriminação e que não há respostas dadas ao acaso, ficando como parâmetro a ser avaliado somente a dificuldade dos itens; 2) o modelo logístico de dois parâmetros (Birnbaum, 1968), que avalia a dificuldade e a discriminação dos itens, assumindo que não hajam respostas dadas ao acaso; e 3) o modelo de três parâmetros de Lord (1980), no qual os três parâmetros dos itens são avaliados.

Exemplificando com o modelo de Lord: os valores Θ são expressos em coordenadas cartesianas, tendo na ordenada a probabilidade de resposta correta, isto é, o P,(Θ), e na abeissa o traço latente, o próprio Θ. Este procedimento produz, para cada item, uma ogiva, chamada de curva característica do item *(item characteristic curve* ou ICC), como na Figura 3-3.
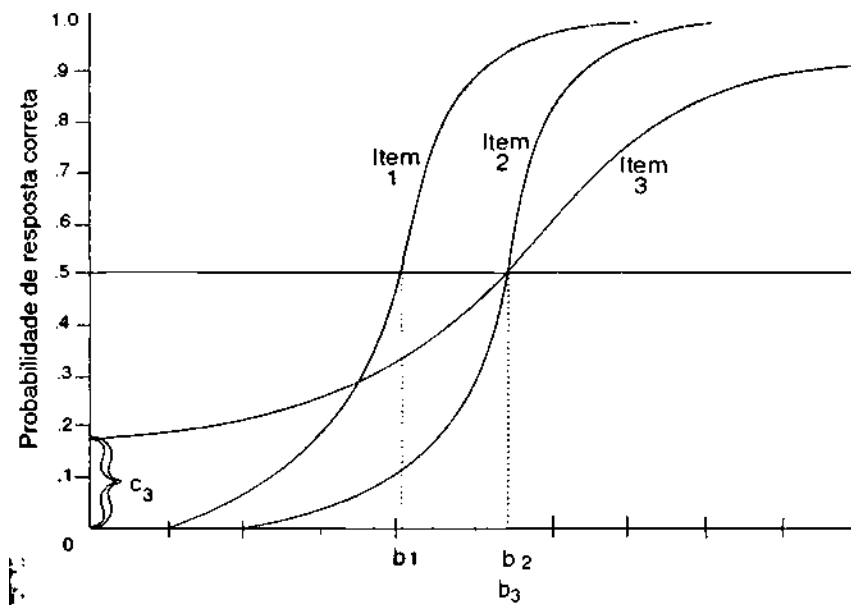
Fig. 3-3 — ICC para três itens.

Na ilustração da Figura 3-3, os três parâmetros aparecem nas seguintes posições: o a é representado pela inclinação da curva na altura do ponto de inflexão, isto é, onde a curva corta a linha que representa a

probabilidade .50 de resposta correta (50%); quanto mais íngreme esta curvatura, mais discriminativo é o item. O b é representado pela distância na linha dos X (abcissa) que corresponde ao ponto determinado pela perpendicular que vem do ponto de inflexão da curva. O c é definido pela assíntota inferior da curva; quando esta assíntota não atinge a abcissa, há respostas dadas ao acaso, e o tamanho destas respostas é definido pela distância que vai do ponto 0, na abcissa, até o ponto onde a curva corta a ordenada.

A adequação do uso da IRT depende de uma suposição que ela própria não pode verificar, a saber, a unidimensionalidade do teste, isto é, a homogeneidade do conjunto de itens que supostamente devem estar medindo um mesmo traço. Este problema é bastante grave, dado que ele atinge algo de crucial na psicometria — que é a questão da representação comportamental do traço latente — e incide diretamente sobre a questão básica da validade dos testes. Tipicamente, a dimensionalidade do teste é verificada através da análise fatorial, e somente após ela seria justificado o uso de uma análise de tipo IRT. Isto, evidentemente, vale também para a análise dos itens baseada no escore total. A IRT tem uma vantagem sobre a última técnica, porque, pelo menos, analisa os parâmetros do item independentemente um do outro. O que fica de problemático neste contexto é que a análise fatorial se fundamenta em equações lineares e, ao que parece, as correlações entre os itens seriam normalmente de caráter curvilinear. Conseqüentemente, a prova da unidimensionalidade do teste, que mede um e um só traço, fica difícil de ser cabalmente demonstrada.

4) Vieses de resposta

Independentemente da qualidade dos itens, a resposta aos mesmos pode ser desvirtuada por fatores relativos ao sujeito que a eles reage. Estes vieses na resposta falseiam os dados, introduzindo correlações espúrias, mesmo em se tratando de bons instrumentos psicológicos. Podemos classificar estes erros em três categorias, em termos de suas causas: cultura/nível socioeconómico, resposta aleatória e resposta estereotipada.

a) Erros associados à cultura

A causa de erros de resposta associados à cultura se relaciona ao problema da transferência de instrumentos psicológicos para outras populações para as quais eles não foram especificamente construídos e validados. E o caso da utilização destes instrumentos para minorias e o da adaptação dos mesmos a outras culturas (tradução de testes).

O problema do uso dos testes com minorias tem recebido grande atenção nos Estados Unidos, sobretudo com a minoria negra. A IRT também vem se preocupando com esta questão no contexto do uso de instrumentos para estudos transculturais (Hambleton, 1991). O problema

que se observa ali é sobretudo a dificuldade relativa de certos itens para grupos de indivíduos com tradições culturais e de experiência diferentes das dos grupos para os quais os testes foram elaborados. Isto significa que pessoas de habilidades similares num dado construto psicológico, mas de culturas diferentes, apresentam diferentes probabilidades de êxito.

Vários métodos estatísticos foram apresentados para lidar com este problema, chamado viés do teste *{test bias),* salientando-se o enfoque que analisa a proporção dos sujeitos que respondem corretamente o item em cada grupo (Angoff, 1982) e a IRT (Ironson, 1982; Hambleton, 1991; Ellis, 1991; Jackson, 1991).

A técnica de Angoff (1982) consiste em transformar as porcentagens de acertos nas duas populações em valores delta e plotá-los em coordenadas cartesianas. No caso de as populações serem similares, esperam-se altas correlações entre as respostas dos sujeitos de ambas as amostras, isto é, os itens se apresentam com dificuldades similares, resultando em um agrupamento dos itens em cima ou ao longo da linha de 45° que passa pelo ponto de origem das coordenadas. Quando estas populações, contudo, forem culturalmente diferentes, o índice geral de dificuldade dos itens pode aparecer mais forte numa que na outra. Neste caso, os valores deltas não se agrupam junto à linha de 45°, mas os pontos (que definem os itens) aparecem mais afastados desta linha, isto é, aparece uma série de itens longe da linha, parecendo estranhos aos demais *(outsiders).* Estes itens são mais difíceis para a amostra de sujeitos onde eles aparecem plotados (vide Figura 3-4).

O problema com esta análise consiste no fato de que itens bem discriminativos tendem a se mostrar *outsiders* e, com isso, correm o perigo de serem eliminados como desviantes. Este problema é evitado com o uso do enfoque da IRT para analisar os itens em termos de desviantes quando aplicados a uma população culturalmente diferente da original, dado que ela analisa os parâmetros de cada item independentemente uns dos outros.

Esta técnica permite analisar a equivalência dos itens quando aplicados a populações culturalmente distintas, identificando os itens que não apresentam tal equivalência, isto é, ¡tens que apresentam um funcionamento diferencial — *differential item functioning* - DIF — (Ellis, 1991). AIRT produz os parâmetros de dificuldade (b) e de discriminação (a) dos ¡tens para as duas amostras culturalmente diferentes, e estes parâmetros podem ser comparados para verificar se são ou não estatisticamente equivalentes ou diferentes. Esta hipótese é testada através do qui-quadrado de Lord (1980; Hulin et al., I983).

Ao se piotar os índices de dificuldade (b) assim calculados das duas amostras em coordenadas cartesianas, os b se alinham ao longo de uma linha paralela à linha de 45°, que não passa pela origem das coordenadas,
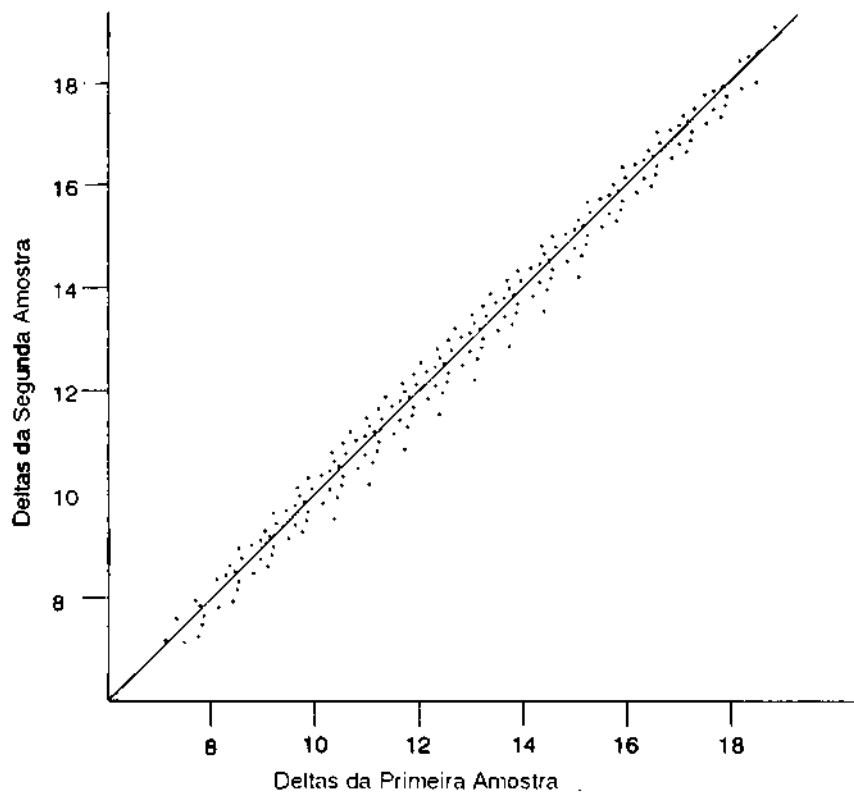
Fig. 3-4 — Distribuição hipotética da dificuldade dos itens em amostras de culturas diferentes.

mas corta, por exemplo, o eixo dos X, indicando que o teste como um todo é mais difícil para a amostra indicada neste eixo (vide Figura 3-5). Os itens mais difíceis para uma ou outra amostra aparecem indentados em direção ao eixo da amostra para a qual tais itens são particularmente difíceis. Assim, por exemplo, o item # 1 é mais difícil para a amostra A, sendo o item #2 mais difícil para a amostra B.

Uma grande vantagem do enfoque da IRT neste particular consiste em possibilitar a descoberta das causas de desvio de certos ¡tens (o fato de serem mais difíceis para uma amostra do que para outra) e, assim, se poder corrigir esta causa e tentar recuperar o item para o teste (Ellis, 1991).

b) A resposta ao acaso

Os fatores que determinam a resposta ao acaso não são determináveis, pois são, por definição, aleatórios. Tal ocorrência pode ser devido a
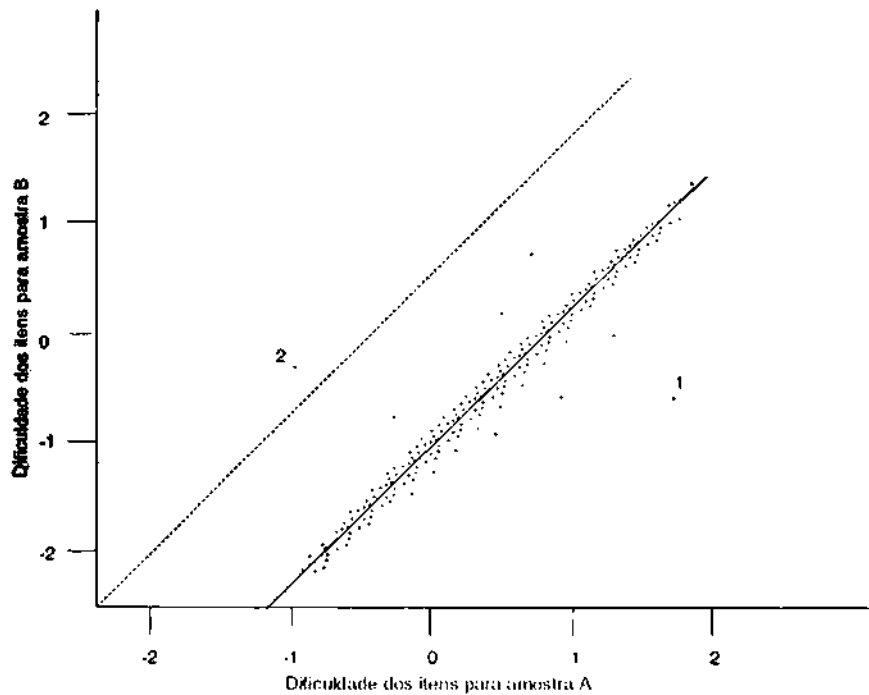
Fig. 3-5 — Distribuição dos b de duas amostras.

inúmeros fatores não sistemáticos, como a má disposição do sujeito em responder ao teste, incompreensão das instruções, gozação e outros. A IRT identifica este tipo de resposta através do parâmetro c.

c) A resposta estereotipada

Trata-se realmente de erros ou de respostas tendenciosas devidos a peculiaridades do sujeito que responde, sobretudo ocorrendo em testes de personalidade e de atitude — são devidos a uma estereotipia na resposta. Dois tipos aparecem salientes: a desejabilidade social e as respostas sistemáticas.

A **desejabilidade social** na verdade corresponde a um traço de personalidade, mas afeta negativamente a objetividade nas respostas de auto-relato. Esta questão foi amplamente discutida por Edwards (1957, 1959), Edwards e Walker (1961), Edwards, Diers e Walker (1962) e Edwards e Walsh (1963). Edwards, inclusive, construiu uma escala para avaliar esta tendência nos sujeitos (Heineman, 1952; Messick, Jackson, 1961); êle define a desejabilidade social como "a tendência dos sujeitos em atribuir a si mesmos, em caso de autodescrição, afirmações de personalidade com valores socialmente desejáveis e em rejeitar aquelas com valores social-

mente indesejáveis" (Edwards, 1957, p.vii). Essa atitude não representa uma vontade de falsear os dados, mas é um desejo (inconsciente) de se apresentar bem diante dos outros. O sujeito não procura intencionalmente mentir sobre si mesmo (neste caso, seria mentira), mas o faz sem dar-se conta disso: quer simplesmente aparecer com bons olhos diante dos outros. Esta tendência é tão comum que parece um traço universal do ser humano. Ela é, igualmente, um problema praticamente sempre presente em inventários de personalidade. A maneira de controlar esta tendência tem sido a elaboração de uma escala de desejabilidade e incluí-la no inventário. Assim, um traço a mais é mensurado pelo inventário; mas fica difícil saber o que fazer com tal dado que alerta sobre o fato de que o sujeito pode bem ter utilizado a mesma tática na resposta aos demais traços medidos pelo inventário. Pelo menos, fica o alerta para a interpretação dos resultados do inventário, quando índices elevados de desejabilidade social estão presentes nos respondentes.

A **resposta sistemática,** por outro lado, representa erros de julgamento. Hã uma série frustrante deste tipo de erros de resposta: efeito de halo, leniência, tendência central, contraste, proximidade, e outras. O controle destes erros tem se mostrado ainda bastante falho na utilização de escalas de avaliação.

O efeito de halo foi cunhado por Thorndike (1920) e ocorre quando "um avaliador tende a avaliar um indivíduo de modo semelhante sobre todas as dimensões" (Guilford, 1959, p. 146). Este erro é inversamente proporcional à variância nas respostas (Borman, 1975), acarretando altas correlações entre diferentes fatores (Gillinsky, 1947; Taylor, Hastman, 1956) e reduzidos desvios padrões (Bernardin, Walter, 1977).

O erro de leniência consiste em dizer "apenas coisas boas a respeito de todo o mundo" (Dunnette, 1983). Estatisticamente, esta tendência é definida como "uma mudança significativa na média das avaliações na direção favorável, de uma condição de avaliação para outra" (Sharon, Bartlett, 1969, p.252).

A tendência central ocorre quando um avaliador tende a colocar todos os sujeitos no centro da escala. É uma tendência na qual "avaliadores hesitam proferir julgamentos extremos... e talvez ocorre mais normalmente quando avaliadores não conhecem suficientemente bem os avaliandos" (Guilford, 1954, p.278).

Erro de contraste consiste na tendência das pessoas avaliarem os outros ao oposto do que se avaliam a si mesmas. Os outros se tornam o contraponto da auto-avaliação. Quem é organizado acha todos os outros desleixados (Murray, 1938).

O controle destes e outros erros da resposta se apresenta difícil, dado que eles têm origem na própria personalidade do sujeito que res-

ponde, tratando-se, portanto, de outros traços da própria personalidade. Tem-se inventado maneiras de contornar tais erros, eliminando, por exemplo, o ponto central (neutro) da escala para inviabilizar a tendência central ou eliminar a parte inferior da escala para descaracterizar a leni-ência, mas tais investidas não têm surtido efeitos suficientes e claros e, assim, estas tendências ainda continuam sendo um problema substancial na medida da personalidade e das atitudes.

Validade

Costuma-se definir a validade de um teste dizendo que ele é válido se de fato mede o que supostamente deve medir. Embora esta definição pareça uma tautologia, na verdade ela não é, considerada a teoria psicométrica exposta acima. O que se quer dizer com esta definição é que, ao se medirem os comportamentos (itens), que são a representação do traço latente, está-se medindo o próprio traço latente. Tal suposição é justificada se a representação comportamental for legítima. Esta legitimação somente é possível se existir uma teoria prévia do traço que fundamente que a tal representação comportamental constitui uma hipótese dedutível desta teoria. A validade do teste (este constituindo a hipótese), então, será estabelecida pela testagem empírica da verificação da hipótese — pelo menos, esta é a metodologia científica. Assim, fica muito estranha a prática corrente na psicometria de se agrupar intuitivamente uma série de itens e, *a posteriori,* verificar estatisticamente o que eles estão medindo. A ênfase na formulação da teoria sobre os traços tem sido muito fraca no passado; com a influência da Psicologia Cognitiva, esta ênfase felizmente está voltando ou deverá voltar ao seu devido lugar na psicometria.

A validação da representação comportamental do traço, isto é, do teste, embora constitua o ponto nevrálgico da psicometria, apresenta dificuldades importantes que se situam em três níveis ou momentos do processo de elaboração do instrumento: o nível da teoria, a coleta empírica da informação e a própria análise estatística da informação.

No nível da teoria se concentram talvez as maiores dificuldades. Na verdade, a teoria psicológica se encontra ainda em estado embrionário, destituída quase que totalmente de qualquer nível de axiomatização, resultando disto uma pletora de teorias, muitas vezes até contraditórias. Basta lembrar de teorias como behaviorismo, psicanálise, psicologia existencialista, psicologia dialética e outras, que postulam princípios irredutíveis entre as várias teorias e pouco concatenados dentro de uma mesma teoria ou, então, em número insuficiente para se poder deduzir hipóteses úteis para o conhecimento psicológico. Havendo esta confusão no campo teórico dos construtos, torna-se extremamente difícil para o psicometrista operacionalizar estes mesmos construtos, isto é, formular

hipóteses claras e precisas para testar ou, então, formular hipóteses psicologicamente úteis. Ainda quando a operacionalização for um sucesso, a coleta da informação empírica não é ¡senta de dificuldades, como, por exemplo, a definição inequívoca de grupos critérios onde estes construios possam ser idealmente estudados. Mesmo no nível das análises estatísticas encontramos problemas. Pela lógica da elaboração do instrumento, a verificação da hipótese da legitimidade da representação dos construtos se faz pela análise fatorial (confirmatoria), que procura identificar, nos dados empíricos, os construtos previamente operacionalizados no instrumento. Acontece que a análise fatorial faz algumas postulações fortes que nem sempre se coadunam com a realidade dos fatos. Por exemplo, a análise fatorial assume que as respostas dos sujeitos aos itens do instrumento são determinadas por uma relação linear destes com os traços latentes. Todos os métodos atuais de análise fatorial postulam esta relação linear. Há, ainda, o grave problema da rotação dos eixos, a qual permite a demonstração de um número sem fim de fatores para o mesmo instrumento.

Diante destas dificuldades, os psicometristas recorrem a uma série de técnicas para viabilizar a demonstração da validade dos seus instrumentos. Fundamentalmente, estas técnicas podem ser reduzidas a três grandes classes: técnicas que visam à validade de construto, validade de conteúdo e validade de critério (APA, I954).

Validade de construto

A validade de construto ou de conceito é considerada a forma mais fundamental de validade dos instrumentos psicológicos — e com toda a razão, dado que ela constitui a maneira direta de verificar a hipótese da legitimidade da representação comportamental dos traços latentes e, portanto, se coaduna exatamente com a teoria psicométrica. Historicamente, o conceito de construto entrou na psicometria através da APA Committee on Psychological Tests, que trabalhou de 1950 a 1954 e cujos resultados se tornaram as recomendações técnicas para os testes psicológicos (APA, 1954).

O conceito de validade de construto foi elaborado com o já clássico artigo de Cronbach e Meehl (1955) "Construct validity in psychological tests", embora o conceito já tivesse uma história sob outros nomes, tais como validade intrínseca, validade fatorial e até validade aparente *(face validity).* Estas várias terminologias demonstram a confusa noção que construto possuía. Embora tenham tentado clarear o conceito de validade de construto, Cronbach e Meehl ainda o definem como a característica de um teste enquanto mensuração de um atributo ou qualidade, o qual não tenha sido "definido operacionalmente". Reconhecem, entretanto, que a validade de construto reclama por um novo enfoque científico. De fato, definir esta validade do modo que eles definiram parece um pouco estranho em ciência, dado que conceitos não definidos operacionalmente

não são suscetíveis de conhecimento científico. Conceitos ou construios são cientificamente pesquisáveis somente se forem, pelo menos, passíveis de representação comportamental adequada; do contrário, serão conceitos metafísicos e não científicos. O problema é que os autores, sintetizando, aliás, a atitude geral dos psicometristas da época, para definir validade de construto, partiram do teste, isto é, da representação comportamental, em vez de partir da teoria psicométrica que se fundamenta na elaboração da teoria do construto (dos traços latentes). O problema não é descobrir o construto a partir de uma representação existente (teste), mas sim descobrir se a representação (teste) constitui uma representação legítima, adequada do construto. Este enfoque exige uma colaboração, bem mais estreita do que existe, entre psicometristas e Psicologia Cognitiva.

A validade de construto de um teste pode ser trabalhada sob dois ângulos: a análise da representação comportamental do construto e a análise por hipótese.

1) A análise da representação

São utilizadas duas técnicas como demonstração da conformidade da representação do construto: a análise fatorial e a análise da consistência interna.

A análise da consistência interna consiste essencialmente em verificar a homogeneidade dos itens que compõem o teste. Assim, o escore total no teste se torna o critério de decisão, e a correlação entre cada item e este escore total decide a qualidade do item: sendo alta a correlação, o item é retido. O índice *alpha* (a) de Cronbach é tipicamente utilizado como indicador sumário da consistência interna do teste e, conseqüentemente, dos itens que o compõem.

Há alguns problemas com esta técnica como demonstração da conformidade da representação do construto. Primeiramente, o escore total constitui uma dificuldade, dado que ele somente faz sentido se o teste já é, a *priori,* homogêneo. Assim, a correlação de cada item com o escore total já pressupõe que os itens são somáveis, isto é, homogêneos; em outras palavras, se pressupõe que todos os itens sejam uma representação do mesmo traço (unidimensionalidade). Por outro lado, a intercorre-lação entre os itens não é uma demonstração de que estes estejam medindo um e mesmo construto. Suponha a situação de três itens saturados em três fatores, como segue:

	F_1	F_2	F_3
1	**.80**	**.30**	**.30**
2	**.30**	**.80**	**.30**
3	**.30**	**.30**	**.80**

As correlações entre os três itens são todas de 0,57, altas e significativas, mas nem por isso se pode dizer que os três itens estejam medindo uma e a mesma coisa. Na verdade, o item 1 mede especificamente o fator 1 e os outros itens medem outros fatores. Conseqüentemente, a análise da consistência interna dos itens não parece garantir que eles sejam uma representação unidimensional de um construto.

A análise fatorial tem como lógica precisamente verificar quantos construtos comuns são necessários para explicar as covariancias (as intercorrelações) dos itens. As correlações entre os itens são explicadas, pela análise fatorial, como resultantes de variáveis-fonte que seriam as causas destas covariancias. Estas variáveis-fonte são os construtos ou traços latentes de que fala a psicometria. A análise fatorial também postula que um número menor de traços latentes (variáveis-fonte) é suficiente para explicar um número maior de variáveis observadas (itens). Assim, se, a partir de uma teoria, foi construído um teste que mede um único traço latente, a análise fatorial poderia verificar esta hipótese, demonstrando se o tal teste de fato mede um único traço, isto é, que a matriz de intercorrelações entre os itens pode ser reduzida ou explicada adequadamente por um únicofator. Neste caso, temos uma demonstração empírica da hipótese de que o teste constitui uma representação legítima e adequada do construto em questão, isto é, o teste tem validade de construto.

Infelizmente, as coisas não aparecem assim tão nítidas. A razão principal desta dúvida se fundamenta na teoria matemática da análise fatorial que define a relação entre as variáveis observadas e as variáveis-fonte somente em termos de equações lineares. Parece difícil se admitir que as intercorrelações entre os itens possam ser todas elas reduzidas a equações lineares, particularmente quando em, quiçá, nenhum campo da Psicologia e das ciências sociais e do comportamento em geral se encontram tais equações. Encontram-se, sim, equações logarítmicas, exponenciais e outras, isto é, equações não-lineares, como, por exemplo, nas leis da psicofisica e da análise experimental do comportamento.

2) Análise por hipótese

Esta análise se fundamenta no poder de um teste psicológico ser capaz de discriminar ou predizer um critério externo a ele mesmo; por exemplo, discriminar grupos-critério que difiram especificamente no traço que o teste mede. Este critério é procurado de várias formas, havendo quatro entre as mais salientes e normalmente utilizadas: validação convergente-discriminante, idade, outros testes do mesmo construto e a experimentação.

A técnica da **validação convergente-discriminante** (Campbell, Fiske, 1967) parte do princípio de que para demonstrar a validade de construto de um teste é preciso determinar duas coisas: 1) o teste deve

correlacionar significativamente com outras variáveis com as quais o cons-truto medido pelo teste deveria, pela teoria, estar relacionado (validade convergente); e 2) não se correlacionar com variáveis com as quais ele teoricamente deveria diferir (validade discriminante).

Campbell e Fiske (1967, p. 125) apresentam o exemplo da Tabela3-3, a seguir:

Tabela 3-3 — Matriz sintética de Multitraço- -Multimétodo

		Método 1			Método 2			Método 3		
	Traço	A_1	B_1	C_1	A_2	B_2	C_2	A_3	B_3	C_3
Método 1	A_1	(.89)								
	B_1	.51	(.89)							
	C_1	.38	.37	(.76)						
Método 2	A_2	**.57**	.22	.09	(.93)					
	B_2	.22	**.57**	.10	.68	(.94)				
	C_2	.11	.11	**.46**	.59	.58	(.84)			
Método 3	A_3	**.56**	.22	.11	**.67**	.42	.33	(.94)		
	B_3	.23	**.58**	.12	.43	**.66**	.34	.67	(.92)	
	C_3	.11	.11	**.45**	.34	.32	**.58**	.58	.60	(.85)

A ilustração apresenta seis blocos de resultados: três triângulos (com linhas inteiras) e três retângulos (com triângulos de linhas pontilha-das). As diagonais dos blocos-retângulo representam as correlações entre as variáveis medidas por diferentes métodos e contêm a diagonal da validade (convergente); estes valores devem ser altos para mostrar validade de construto. Os valores fora destas diagonais nestes mesmos blocos (os triângulos de linhas pontilhadas) representam as correlações entre diferentes variáveis medidas por diferentes métodos; estes valores devem ser pequenos para mostrar validade de construto (validade discriminante). O mesmo deve ocorrer com as correlações fora das diagonais nos blocos-triângulo (com linhas inteiras), que representam os coeficientes entre variáveis diferentes medidas pelo mesmo método (nas diagonais estão os

coeficientes de precisão). Este método funciona se os métodos e as variáveis diferem o suficiente (máximamente) entre si.

A **idade** é utilizada como critério para a validação de construto de um teste quando este mede traços que são intrinsecamente dependentes de mudanças no desenvolvimento cognitivo/afetivo dos indivíduos, como é o caso, por exemplo, da teoria piagetiana do desenvolvimento dos processos cognitivos e da teoria de Spearman sobre a inteligência. A hipótese a ser testada neste método é a de que o teste que mede o traço X, o qual muda claramente com a idade, é capaz de discriminar distintamente grupos de idades diferentes. O problema com este método consiste no fato de que a maturação psicológica pode assumir dimensões e conotações muito distintas em culturas diferentes, por um lado; por outro, outras variáveis que não o traço em questão podem estar dependentes desta maturação, dificultando ou impossibilitando a definição dos grupos-critério somente em função da idade. Assim, se outras variáveis variam com a idade, pode bem ser que estas sejam as responsáveis pelas mudanças no traço e não a idade especificamente. Isto não seria grave problema se estas outras variáveis variassem do mesmo modo em qualquer contexto cultural ou socioeconòmico, o que obviamente é difícil de assumir. Dentro de uma mesma cultura, o método se apresenta como importante para a determinação da validade de construto.

A **correlação com outros testes** que meçam o mesmo traço é também utilizada como demonstração da validade de construto. O argumento é de que, se um teste X mede validamente o traço Z e o novo teste N se correlaciona altamente com o teste X, então o novo teste mede o mesmo traço medido por aquele teste. O problema com esta técnica consiste no fato de que normalmente um teste de um traço qualquer não se apresenta com tal pureza a se poder afirmar que ele mede exclusivamente o tal traço. De fato, ele mede o traço em termos de um certo nível de covariancia: por exemplo, existe uma correlação de 0,70 entre o teste e o traço, o que eqüivale a uma comunalidade de 49%. Agora, o novo teste correlaciona, digamos, também 0,70 com aquele teste, havendo, portanto, comunalidade de 49% entre os dois testes. Qual será, neste caso, a comunalidade do novo teste com o traço em si? Por azar, poderia acontecer que a comunalidade de 49% entre os dois testes ocorra precisamente com os 51% do primeiro teste que não covariam com o traço; neste caso, a comunalidade do novo teste com o traço seria 0%, isto é, o novo teste seria uma representação totalmente equivocada do traço.

O uso da **intervenção experimental** aparece como logicamente uma das melhores técnicas para se decidir a validade de construto de um teste. Esta técnica consiste em verificar se o teste discrimina claramente grupos-critério 'produzidos' experimentalmente em termos do traço objeto de medida do teste. Assim, um teste que mede ansiedade teria validade de

construto (ansiedade) se discriminasse grupo não-ansioso de grupo ansioso, definidos estes grupos em termos de manipulações experimentais — o ansioso, por exemplo, criado assim através de experiencias provocadoras de ansiedade. Uma vez que se possa garantir que as manipulações feitas nos grupos-criterio atingem exclusivamente o traço em questão, a testagem da hipótese é válida; como, normalmente, estas manipulações supostamente de uma variável de fato afeta uma série de outras variáveis, sobretudo se as variáveis interagirem, fica muito confusa a decisão sobre em que especificamente os grupos-critério diferem e, conseqüentemente, fica inconclusiva a decisão sobre a hipótese de que o teste discrimina os grupos-critério exclusivamente em termos do traço que ele pretende medir.

Em conclusão, a técnica da validação de construto via hipótese, que, de um ponto de vista da metodologia científica, se apresenta como a mais direta e óbvia, esbarra na dificuldade que existe na definição inequívoca do critério a ser utilizado como representante da manifestação do traço.

Deve-se, na verdade, concluir que todas estas técnicas de validação apresentam dificuldades graves, mas nem por isso se justifica o simples abandono das mesmas. Primeiramente, porque em ciência empírica nada existe de perfeito e isento de erro e, em segundo lugar, a consciência destas dificuldades deve servir para melhorar e não abandonar as técnicas. Aliás, é recomendável o uso de mais de uma das técnicas acima analisadas para demonstrar a validade de construto do teste, dado que a convergência de resultados das várias técnicas constitui garantia para a validade do instrumento.

Validade de critério

Concebe-se como validade de critério de um teste o grau de eficácia que ele tem em predizer um desempenho específico de um sujeito. O desempenho do sujeito torna-se, assim, o critério contra o qual a medida obtida pelo teste é avaliada. Evidentemente, o desempenho do sujeito deve ser medido/avaliado através de técnicas que são independentes do próprio teste que se quer validar.

Costuma-se distinguir dois tipos de validade de critério: 1) validade preditiva e 2) validade concorrente. A diferença fundamental entre os dois tipos é basicamente uma questão do tempo que ocorre entre a coleta da informação pelo teste a ser validado e a coleta da informação sobre o critério. Se estas coletas forem simultâneas (mais ou menos), a validação será de tipo concorrente; caso os dados sobre o critério sejam coletados após a coleta da informação sobre o teste, fala-se em validade preditiva. O fato de a informação ser obtida simultaneamente ou posteriormente à do próprio teste não é um fator tecnicamente relevante à validade do teste; relevante, sim, é a determinação de um critério válido. Aqui se situa

precisamente a natureza central deste tipo de validação dos testes: 1) definir um critério adequado e 2) medir, válida e independentemente do próprio teste, este critério.

Quanto à adequação dos critérios, pode-se afirmar que há uma série destes que são normalmente utilizados, quais sejam:

1) Desempenho acadêmico

Talvez seja o critério mais utilizado na validação de testes de inteligência. Consiste na obtenção do nível de desempenho escolar dos alunos, seja através das notas dadas pelos professores, seja pela média acadêmica geral do aluno, seja pelas honrarías acadêmicas que o aluno recebeu, ou seja mesmo pela avaliação puramente subjetiva dos alunos em termos de "inteligente" por parte dos professores ou colegas. Embora seja amplamente utilizado, este critério tem igualmente sido amplamente criticado, não em si mesmo, mas pela deficiência que ocorre na sua avaliação. É sobejamente sabida a tendenciosidade por parte dos professores em atribuir as notas aos alunos, tendenciosidade nem sempre consciente, mas decorrente de suas atitudes e simpatias em relação a este ou aquele aluno. Esta dificuldade poderia ser sanada até com certa facilidade, se os professores tivessem o costume de aplicar testes de rendimento que possuíssem validade de conteúdo, por exemplo. Como esta tarefa é dispendiosa, o professor tipicamente não se dá ao trabalho de validar (validade de conteúdo) suas provas acadêmicas.

Neste contexto, é também utilizado como critério de desempenho acadêmico o nível escolar do sujeito: sujeitos mais avançados, repetentes e evadidos — a suposição sendo de que quem continua regularmente ou está avançado em termos de sua idade possui mais habilidade. Evidentemente, nesta história não entra somente a questão da habilidade, mas muitos outros fatores sociais, de personalidade, etc, tornando este critério bastante ambíguo e espúrio.

2) Desempenho em treinamento especializado

Trata-se do desempenho obtido em cursos de treinamento em situações específicas, como no caso de músicos, pilotos, atividades mecânicas ou eletrônicas especializadas, etc. No final deste treinamento há tipicamente uma avaliação, a qual produz dados úteis para servirem de critério de desempenho do aluno.

3) Desempenho profissional

Trata-se, neste caso, de comparar os resultados do teste com o sucesso/fracasso ou o nível de qualidade do sucesso dos sujeitos na própria situação de trabalho. Assim, um teste de habilidade mecânica pode ser testado contra a qualidade de desempenho mecânico dos sujeitos na

oficina de trabalho. Evidentemente continua a dificuldade de levantar adequadamente a qualidade deste desempenho em serviço dos sujeitos.

4) Diagnóstico psiquiátrico

Muito utilizado para validar testes de personalidade/psiquiátricos. Os grupos-critério são aqui formados em termos da avaliação psiquiátrica: normais vs. neuróticos, psicópatas vs. depressivos, etc. Novamente, a dificuldade continua sendo a adequação das avaliações psiquiátricas feitas pelos psiquiatras.

5) Diagnóstico subjetivo

Avaliações feitas por colegas e amigos podem servir de base para estabelecer grupos-critério. É utilizada esta técnica sobretudo em testes de personalidade, onde é difícil encontrar avaliações mais objetivas. Assim, os sujeitos avaliam seus colegas em categorias ou dão escores em traços de personalidade (agressividade, cooperação, etc), baseados na convivência que eles têm com os colegas. Nem precisa mencionar as dificuldades enormes que tais avaliações apresentam em termos de objetividade; contudo, a utilização de um grande número de juizes poderá diminuir os vieses subjetivos nestas avaliações.

6) Outros testes disponíveis

Os resultados obtidos através de outro teste válido que meça o mesmo construto que o teste a ser validado servem de critério para determinar a validade do novo teste. Aqui fica a pergunta óbvia: para que criar outro teste, se já existe um que mede validamente o que se quer medir? A resposta se baseia numa questão de economia: utilizar um teste que demanda muito tempo para ser respondido ou apurado como critério para validar um teste que gaste menos tempo.

Todos estes critérios podem ser considerados bons e úteis para fins de validação de critério. A grande dificuldade em quase todos eles se situa na demonstração da adequação da medida deles: em geral, a medida dos mesmos é precária, deixando, por isso, muita dúvida quanto ao processo de validação do teste. Entretanto, há exemplos famosos de testes validados através deste método, como é o caso do MMPI.

Validade de conteúdo

Um teste tem validade de conteúdo se ele constitui uma amostra representativa de um universo finito de comportamentos (domínio); é aplicável quando se pode delimitar a *priorie* claramente um universo de comportamentos, como é o caso em testes de desempenho, que pretendem cobrir um conteúdo delimitado por um curso específico.

Para viabilizar um teste com validade de conteúdo, é preciso que se façam as especificações do teste antes da construção dos itens. Estas

especificações comportam a definição de três grandes temas: 1) definição do conteúdo, 2) explicitação dos processos psicológicos (os objetivos) a serem avaliados e 3) determinação da proporção relativa de representação no teste de cada tópico do conteúdo.

Quanto ao conteúdo, trata-se de detalhá-lo em termos de tópicos e subtópicos e de explicitar a importância relativa de cada tópico dentro do teste. Tais procedimentos evitam a super-representação indevida de alguns tópicos e sub-representação de outros por vieses e pendores pessoais do avaliador. Claro que será sempre o avaliador ou equipe de avaliadores quem vai definir este conteúdo e a relativa importância de suas partes, mas esta definição deve ser tomada antes da construção dos itens, garantindo certa objetividade, pelo menos, nas decisões.

Quanto aos objetivos, um teste não deve ser elaborado para avaliar exclusivamente um processo. Como na aprendizagem entram em ação vários processos psicológicos, há interesse que todos eles sejam avaliados por um teste de conteúdo. Por exemplo, o teste deverá conter itens que avaliam a memória (reproduzir), a compreensão (conceituar, definir), a capacidade de comparação (relacionar) e de aplicação dos princípios aprendidos (solucionar problemas, transferência da aprendizagem).

Para facilitar a especificação do teste, pode-se utilizar uma tabela de dupla entrada, com o detalhamento dos objetivos (processos) no topo e o detalhamento dos tópicos à esquerda, explicitando, no corpo da tabela, o número de itens, conforme Tabela 3-4.

Tabela 3-4 — Especificação do conteúdo e processos para uma prova de rendimento em estatística descritiva

Conteúdo	Processos (objetivos)			Total
	Conceituar	Relacionar	Aplicar	
Freqüência	2	3	1	6
Tendência Central	3	1	5	9
Variabilidade	3	2	4	9
Total	8	6	10	24

Precisão (fidedignidade)

O problema que se enquadra sob o conceito de fidedignidade vem relatado sob uma série de outras denominações, como precisão, fidedignidade, constância, consistência interna, confiabilidade, estabilidade, confiança, homogeneidade. As mais genéricas e, por isso, as mais utilizadas são precisão e fidedignidade.

Estas diferentes designações mostram a variabilidade de conceitos que precisão assume, dependendo do aspecto que este parâmetro quer salientar do teste. Na verdade, fidedignidade cobre aspectos diferentes de um teste, mas todos eles se referem a quanto os escores de um sujeito se mantêm idênticos em ocasiões diferentes; por exemplo, os escores obtidos num tempo 1 e num tempo 2 para os mesmos sujeitos. Esta ocorrência (identidade dos escores) evidentemente supõe que o traço que o teste mede se mantenha constante sob estas diferentes ocasiões, como é suposto ser o caso, por exemplo, na maioria dos traços de personalidade e de aptidão. Não seria o caso num teste de humor, porque este traço por natureza varia de momento para outro, e um teste válido de humor produziria escores necessariamente diferentes. Assim, o conceito de fidedignidade, na verdade, se refere ao quanto o escore obtido no teste se aproxima do escore verdadeiro do sujeito num traço qualquer; isto é, a fidedignidade de um teste está intimamente ligada ao conceito da variância erro, sendo este definido como a variabilidade nos escores produzida por fatores estranhos ao construto. Aparece, assim, claro que a fidedignidade de um teste depende da questão do erro da medida, especificamente do erro produzido pelo próprio instrumento: quanto o escore produzido pelo teste se distancia do escore verdadeiro do sujeito no traço em questão, isto é, a valor theta individual na IRT.

Para melhor conceber esta problemática, é preciso se referir à variância verdadeira e variância erro. Um procedimento de medida qualquer, por exemplo, os escores em um teste, produz uma variabilidade nos resultados que, em parte, é provocada pelas diferenças no próprio traço medido entre diferentes sujeitos, parte pela imprecisão do próprio instrumento e parte, ainda, por uma série de outros fatores aleatórios. A fidedignidade da medida depende do tamanho da variância erro, que é precisamente a variabilidade nos resultados provocada por estes fatores aleatórios e pela imprecisão do instrumento. Expressa mais positivamente, a fidedignidade de um instrumento diz respeito ao montante de variância verdadeira que ele produz *vis-à-vis* a variância erro, isto é, quanto maior a variância verdadeira e menor a variância erro, mais fidedigno o instrumento: um escore preciso é um escore que se aproxima do valor verdadeiro, expresso estatisticamente pelo erro padrão da medida (tratado mais adiante).

A definição estatística da fidedignidade é feita através da correlação entre escores de duas situações produzidos pelo mesmo teste. Se o teste é preciso, esta correlação deve não somente ser significativa, mas se aproximar da unidade (cerca de 0,90). De fato, uma correlação de 0,70, por exemplo, expressaria uma comunalidade de apenas 49% entre as duas situações provocadas pelo mesmo teste nos mesmos sujeitos. Neste caso, a variância comum, digamos a variância verdadeira, seria menor que a variância erro, demonstrando que o teste não produz resultados fidedignos,

isto é, o teste não possui precisão. Esta correlação, no caso do parametro de fidedignidade ou precisão, é referida como o coeficiente de precisão ou de fidedignidade.

Dependendo da técnica utilizada para demonstração da precisão de um teste, surgem vários tipos de precisão: teste-reteste, formas paralelas, consistência interna.

Precisão teste-reteste

Este tipo de precisão consiste em calcular a correlação entre as distribuições de escores obtidos num mesmo teste pelos mesmos sujeitos em duas ocasiões diferentes de tempo. A correlação de 1,00 seria obtida se não houvesse variância erro provocada pelo teste ou outros fatores aleatórios, como fatores não controlados nos sujeitos ou na situação de testagem. Quanto mais longo o período de tempo entre a primeira e a segunda testagem, mais chances haverá de fatores aleatórios ocorrerem, diminuindo o coeficiente de precisão. Este intervalo de tempo permite a ação dos fatores mencionados por Campbell e Stanley (1%3) sob o tema de fontes de erro devido à história, maturação, retestagem e às interações entre estes fatores, bem como ao próprio instrumento. Por isso, vêem-se as graves dificuldades que apresenta este tipo de análise da fidedignidade de um teste; particularmente grave aparece aqui a questão da maturação, isto é, se o próprio traço matura (se desenvolve, modifica), esta análise da precisão torna-se errônea, dada sobretudo a eventualidade de que a maturação do traço se processe diferencial mente para os diversos sujeitos testados. Além disso, e particularmente em testes de aptidão, a testagem constitui um treinamento, e provavelmente diferencial, para os sujeitos, o que provocará diferenças na retestagem entre os mesmos, reduzindo novamente o coeficiente de precisão do teste. Para contornar estas dificuldades, outros tipos de análises foram elaboradas, como a das formas alternativas ou análise da consistência interna.

Precisão de formas alternativas

Neste caso, os sujeitos respondem a duas formas paralelas do mesmo teste, e a correlação entre as duas distribuições de escores constitui o coeficiente de precisão do teste. A condição necessária para que esta análise seja válida se situa na demonstração de que as amostras de conteúdo (de itens) em ambas as formas sejam equivalentes, isto é, que os itens possuam níveis equivalentes de dificuldade e de discriminação em ambas. Estes parâmetros podem ser facilmente verificados através da IRT; há, contudo, algumas dificuldades neste tipo de análise: as duas formas são aplicadas em sucessão imediata, não eliminando assim totalmente o efeito do intervalo de tempo, resultando na possível introdução de efeitos da história e do treinamento (prática) obtido ao responder à primeira das formas alternativas; aparece facilmente um efeito repetitório, dado que os

itens de ambas as formas são similares, produzindo efeitos motivacionais negativos no responderte. Além disso, não é tarefa fácil construir formas alternativas, quando a construção de um só teste já é uma tarefa dispendiosa, razão pela qual poucos testes aparecem no mercado com formas alternativas.

Precisão da consistência interna

As várias técnicas de estabelecer este tipo de precisão visam verificar a homogeneidade da amostra de itens do teste, ou seja, a consistência interna do teste. As técnicas mais utilizadas são: duas metades, Kuder-Richardson e alfa de Cronbach. Todas elas exigem aplicação do teste em apenas uma única ocasião, evitando totalmente a questão da constância temporal.

1) Precisão das duas metades

Os sujeitos respondem a um único teste numa única ocasião. O teste é dividido em duas partes equivalentes, e a correlação é calculada entre os escores obtidos nas duas metades. Nao é importante como o teste é dividido em duas metades, conquanto que estas sejam equivalentes. Na prática, contudo, as duas formas mais normalmente utilizadas são a divisão do teste em primeira metade e segunda metade ou em itens pares e itens ímpares. Para efetuar esta análise, de fato o teste não precisa ser homogêneo, isto é, no qual todos os itens medem o mesmo traço (por exemplo, itens somente verbais ou numéricos); o que é fundamental é que as duas metades emparelhem itens homogêneos: verbal com verbal, numérico com numérico, etc.

Neste tipo de precisão, é preciso notar que o cálculo da correlação se baseia somente na metade do teste. Assim, num teste de 100 itens, a correlação se basearia somente em 50 itens. Como o número de itens afeta o tamanho do coeficiente de correlação, é preciso corrigir este coeficiente para que leve em consideração a extensão total do teste e, assim, produzir um coeficiente de precisão mais justo para o teste. Esta correção é feita através da fórmula de Spearman-Brown:

$$r_{tt} = \frac{nr_{12}}{1 + r_{12}},$$

onde, r_{tt} é o coeficiente de precisão calculado, r_{12} é o coeficiente de correlação entre as duas metades do teste e n é o número de vezes em que o teste foi dividido. Assim, um teste dividido em duas metades, o n será 2, porque ele deve ser aumentado 2 vezes para se obter a forma total do teste.

2) Precisão de Kuder-Richardson

Esta técnica, criada por Kuder e Richardson (1937), se baseia na análise de cada item individual do teste. Os autores desenvolveram várias fórmulas sendo a mais utilizada a fórmula 20, que segue:

$$r_{tt} = (\frac{n}{n-1})\frac{DP_t^2 - \sum pq}{DP_t^2},$$

onde, r„ é o coeficiente de precisão do teste, n o número de itens do teste, DP,' o desvio padrão dos escores totais do teste e Σpq é o somatório do produto da proporção de sujeitos que passaram (p) e dos que não passaram (q) cada item.

Cronbach (1951) mostrou que esta técnica produz um coeficiente de precisão do teste que corresponde à média dos coeficientes de todas as metades em que o teste possa ser dividido, mas somente quando se utiliza a fórmula de Rulon (1939), que trabalha com as variâncias das diferenças entre as duas metades, e não a simples correlação com a correção de Spearman-Brown, segundo observaram Novick e Lewis (1967). Esta equivalência de coeficientes, contudo, ocorre em testes homogêneos, porque nos testes heterogêneos os coeficientes de Kuder-Richardson são normalmente menores, dado que esta técnica não trabalha com diferenças entre pares de itens e sim com a variância de todos os itens.

3) Alfa de Cronbach

Esta técnica de Cronbach (1951) constitui uma extensão da de Kuder-Richardson. Esta última é aplicável somente quando a resposta ao item é dicotômica — certo e errado, por exemplo. Entretanto, quando a resposta ao item pode assumir mais de duas alternativas, o valor Σpq é substituído por $\Sigma(DP_t^2)$, a soma dos desvios padrões de cada item. Esta fórmula genérica é a seguinte:

onde DP_t^2 é a variância de todo o teste e $\Sigma(DP_t^2)$ é o somatório das variâncias de cada item do teste.

Precisão na apuração dos escores

Existe toda uma série de fatores que podem afetar negativamente os resultados de um teste, tais como distração, cansaço, etc. Para a maioria destes fatores pode-se encontrar alguma maneira de controlá-los experimentalmente através, por exemplo, da padronização das condições de testagem. Há, contudo, um fator importante na apuração dos resultados de testes não-objetivos, onde a opinião do apurador entra como fator de decisão. Neste caso, é preciso que mais de um apurador seja utilizado para se garantir um resultado preciso no teste. A concordância, expressa pela correlação entre as avaliações de diferentes apuradores, produzirá um índice de precisão entre apuradores. Obviamente este índice deve ser positivo e alto para garantir validade dos resultados.

O erro padrão de medida

Para estabelecer a precisão de um teste, pode-se, em lugar de calcular um coeficiente de fidedignidade, calcular o erro provável da medida incorrido pelo teste. Este erro é expresso em termos padrões e é o seguinte:

onde o erro padrão da medida **(EPM)** se expressa em termos do desvio padrão do teste (DP_t) e do coeficiente de precisão do mesmo teste (r_{tt}) obtidos na mesma amostra de sujeitos.

Este índice se apresenta muito útil na interpretação de escores individuais, pois com ele se pode definir os limites do intervalo dentro do qual mais provavelmente se situa o escore verdadeiro do sujeito. Estes limites são expressos pelo desvio padrão: ± 1 **EPM** em torno do escore obtido dá os limites dentro dos quais há uma probabilidade de 68,26% de se situar o escore verdadeiro; se se optar por 2 **EPM,** esta probabilidade sobe para 07,72% (vide curva normal). Assim, se um sujeito obteve um QI de KM) num teste cujo **DP** é 15 e o r_{tt} = .91, seu escore verdadeiro se situa entre os seguintes limites expressos pelo **EPM** de:

$$15\sqrt{1-.91} = 15\sqrt{.09} = 15(.30) = 04.5$$

Neste caso, os limites serão 100 ± 4,5 — ou seja, 104,5 e 95,5.

Existe na literatura uma louvável tendência de se apresentar os resultados dos sujeitos num teste em termos de faixas definidas pelo **DPM** em vez de escores isolados. Esta prática permite, igualmente, comparar com maior precisão a diferença entre dois escores, observando se as faixas dos dois escores se sobrepõem ou não.

Padronização (normas)

Padronização, em seu sentido mais geral, se refere à necessidade de existir uniformidade em todos os procedimentos no uso de um teste válido e preciso: desde as precauções a serem tomadas na aplicação do teste (uniformidade das condições de testagem, controle do grupo, instruções padronizadas e motivar os examinandos pela redução da ansiedade) até o desenvolvimento de parâmetros ou critérios para a interpretação dos resultados obtidos. Em seu sentido mais técnico de parâmetro psicométrico, a padronização se refere a este último aspecto, isto é, como interpretar os resultados.

Um escore bruto produzido por um teste necessita ser contextualizado para poder ser interpretado. Obter, por exemplo, 50 pontos num teste de raciocínio verbal e 40 num de personalidade não oferece nenhuma informação. Mesmo se dissermos que o sujeito acertou 80% das questões

não diz muito, visto que o teste pode ser fácil (80% então seria pouco) ou difícil (80% então seria muito). Na verdade, qualquer escore deve ser referido a algum padrão ou norma para adquirir sentido. Uma tal norma permite situar o escore de um sujeito, permitindo 1) determinar a posição que o sujeito ocupa no traço medido pelo teste que produziu o tal escore e 2) comparar o escore deste sujeito com o escore de qualquer outro sujeito. O critério de referência ou a norma de interpretação é constituído tipicamente por dois padrões: 1) o nível de desenvolvimento do indivíduo humano (normas de desenvolvimento) e 2) um grupo padrão constituído pela população típica para a qual o teste é construído (normas intragrupo).

Normas de desenvolvimento

As normas de interpretação dos escores de um teste baseadas no desenvolvimento se fundamentam no fato do desenvolvimento progressivo (nos vários aspectos de maturação psicomotora, psíquica, etc.) pelo qual o indivíduo humano passa ao longo de sua vida. Neste sentido, são utilizados, como critério de norma, três fatores: idade mental, série escolar e estágio de desenvolvimento.

1) A idade mental

Este critério foi criado por Binet e Simon (1905). Estes autores falavam de nível mental, depois popularizado como idade mental. Binet e Simon separaram empiricamente uma série de 54 questões/tarefas em 11 níveis de idade cronológica: 3 a 10 anos (oito níveis), 12, 15 anos e idade adulta. As questões que eram respondidas corretamente pela média de crianças/sujeitos de uma idade cronológica X definiam o nível/idade mental correspondente a esta idade cronológica. Assim, a um sujeito que respondia a todas as questões que as crianças de 10 anos eram capazes de responder era atribuída a idade mental de 10 anos.

Na adaptação norte-americana da escala de Binet-Simon, a Stanford-Binet (Terman, Merrill, 1960), a idade mental (IM) foi expressa em termos da idade cronológica (IC), resultando no quociente intelectual, o QI, através da fórmula:

Assim, quem responde a todas as questões correspondentes à sua idade cronológica possui um QI de 100 (por exemplo, para uma criança de 10 anos: QI = 100 x (10/10) = 100). A interpretação dos resultados em termos de QI se faz através da Tabela 3-5.

2) Série escolar

Este critério é utilizado para testes de desempenho acadêmico e somente faz sentido quando se trata de disciplinas que são oferecidas numa

Tabela 3-5 — Interpretação dos escores de QI

QI		Interpretação
140-160	....	Definitivamente superior
120-139	....	Superior
110-119	....	Médio Superior
90-109	....	Normal ou médio
80-89	....	Médio inferior
70-79	....	Deficiência limítrofe
50-69	....	Cretino
30-49	....	Imbecil
29	....	Idiota

seqüência de várias séries escolares. As normas são aqui estabelecidas computando-se o escore bruto médio obtido pelos alunos em cada série, resultando num escore típico para cada série. Desta forma, a criança que obtém o escore bruto típico da 4^9 série obtém o escore padronizado de 4.

3) Estágio de desenvolvimento

Este critério é utilizado por pesquisadores na área da psicologia da criança que estudam o desenvolvimento mental e psicomotor em termos de idades sucessivas de desenvolvimento, como Gesell e Piaget.

Gesell e colaboradores (Ames, 1937; Gesell, Amatruda, 1947; Halverson, 1933; Knoblock, Pasamanick, 1974) desenvolveram normas para oito idades típicas (de 4 semanas a 36 meses) de desenvolvimento das crianças nas áreas do comportamento motor, adaptativo, da linguagem e social.

Piaget e seus colaboradores estudaram o desenvolvimento cognitivo e estabeleceram uma seqüência de estágios sucessivos deste desenvolvimento (sensório-motor, pré-operacional, operacional concreto, operacional formal). Seguidores da escola piagetiana desenvolvem testes utilizando estes estágios como critério de interpretação dos escores (Laurendeau, Pinard, 1962, 1970; Pinard, Laurendeau, 1964).

Normas intragrupo

O critério de referência dos escores é o grupo ou a população para a qual o teste foi construído. Aqui o escore do sujeito toma sentido em relação aos escores de todos os sujeitos da população. De fato, ele é referenciado em termos I) do posto percentílico ou 2) do desvio normal (z). Como tipicamente não são conhecidos os escores da população, é sobre uma amostra representativa desta que as normas são estabelecidas.

1) Posto percentílico

O escore do sujeito é expresso em termos de percentil. Este posto indica quanto por cento de todos os sujeitos da população (amostra) estão abaixo dele. Assim, se 40% dos sujeitos obtiveram um escore bruto menor do que 20, este escore será expresso como percentil 40, o que indica que 40% dos sujeitos têm escore menor que 20 e 60% têm escore maior. Um percentil de 50 indica que o sujeito se situa na mediana dos escores da amostra. Usa o intervalo semi-interquartílico (Q) em torno desta mediana para definir o significado relativo dos postos dos sujeitos. Q = (Q3-Ql)/2, onde Ql é o percentil 25 e o Q3 o percentil 75.

Os escores percentílicos são fáceis de calcular e são de compreensão simples. A grande dificuldade da escala percentílica se situa no fato de que as distâncias entre escores sucessivos não são constantes, mas variam segundo a posição do escore estar no início/fim da escala ou no meio dela. De fato, os intervalos entre os percentis medianos são menores do que os dos extremos da escala, como aparece ilustrado na Figura 3-6.

2) Escore padrão

As normas baseadas no escore padrão (escore Z) se fundamentam no cálculo deste escore Z correspondente ao escore bruto do sujeito. Este cálculo é feito de duas formas bastante distintas, que resultarão ou num escore padrão ou num escore padrão normalizado. O primeiro é feito através de uma transformação linear e o segundo através de uma transformação não-linear.

O escore padrão linear é calculado pela fórmula:

$$Z = \frac{X-M}{DP},$$

onde X = escore bruto, M - média do grupo e DP= desvio padrão.

O escore padrão normalizado é calculado através das tabelas da curva normal, e consiste essencialmente em transformar as porcentagens em escores Z, como ilustrado na Figura 3-6.

As duas formas dão resultados idênticos se as distribuições de freqüência forem normais. Quanto mais as distribuições se afastam da normalidade, menos recomendável é a utilização da transformação não-linear dos escores.

De qualquer forma que o Z seja obtido, as normas baseadas nele normalmente utilizam algumas transformações lineares ulteriores para evitar duas dificuldades de uma escala de Z: 1) a presença de escores negativos, pois o Z vai de menos infinito a mais infinito (mais praticamente, de -5 a +5), e 2) a presença de decimais. Para eliminar estas duas dese-legâncias, tipicamente o Z é multiplicado por um coeficiente e ao produto
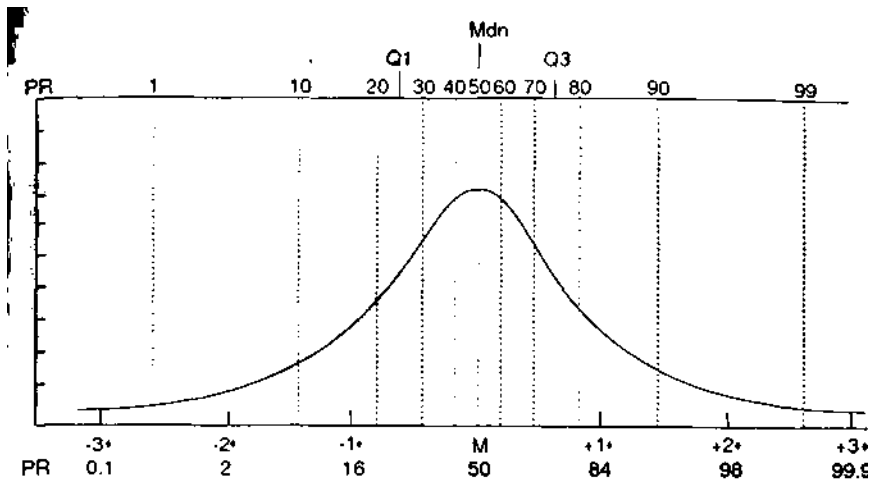
Fig. 3-6 — Distribuição normal e de postos percentílicos.

é agregada uma constante. Tanto o coeficiente de multiplicação do Z quanto a constante somada são arbitrárias, resultando em tantas formas de normas quantas imagináveis. Contudo, alguns desses valores são rotineiramente mais utilizados, produzindo normas já tradicionalmente conhecidas, como: o escore T, os estaninos, o desvio QI, o escore CEEB e vários outros. As fórmulas de transformação para algumas destas normas são:

T = 50 + 10z
Desvio QI = 100 + 15z (Escalas de Wechsler) ou
Desvio QI = 100 + 16z (Stanford-Binet) CEEB =
500 + 100z

CONCLUSÃO GERAL

A psicometria clássica já possui uma longa história e tem conseguido sucessos importantes em nível mundial. Contudo, devido à sua origem de longa data, ela sofre de alguns problemas ou limitações que advem de duas fontes: 1) ela surgiu dentro de um contexto histórico no qual a Psicologia em geral era ditada pelo enfoque positivista e 2) se valeu dos progressos da Estatística na qual predominavam as estatísticas paramétricas e univariadas.

Da primeira fonte surgem as dificuldades referentes ao precário embasamento psicológico teórico que a psicometria ainda sofre. Na verdade, os psicometristas em geral não dão a devida atenção à fundamentação teórica na hora da elaboração dos seus instrumentos psicologi-

cos, donde resultam instrumentos que muitas vêzes não se sabe exatamente o que estão medindo ou, pelo menos, se estão medindo algo de psicologicamente relevante. Este problema pode ser devidamente sanado com o progresso da Psicologia Cognitiva, que dá a devida importância aos componentes dos traços latentes, estudados em seus próprios méritos e não, como vem fazendo a psicometria clássica, como interpretações que se dão às resultantes de análises estatísticas (análise fatorial) feitas sobre um agregado mais ou menos aleatoriamente agrupado de itens ou tarefas respondidos por uma amostra de sujeitos. Da segunda fonte surge uma série de limitações devido ao uso de estatísticas nem sempre adequadas aos dados coletados, como, por exemplo, as análises dos itens baseadas no escore total e a análise da fidedignidade do teste baseada na correlação ou no erro de medida. Estas últimas limitações podem e estão sendo superadas pelo desenvolvimento da nova teoria psicométrica, a *Item Response Theory.*

REFERÊNCIAS BIBLIOGRÁFICAS

AMERICAN PSYCHOLOGICAL ASSOCIATION. Committee on Psychological Tests. Technical recommendations for psychological tests and diagnostic techniques. *Psychological Bulletin Supplement.* Washington, D.C., v.51, n.2. part 2, p.1-38. 1954.

AMES, L.B. The sequential patterning of prone progression in the human infant. *Genetic Psychology Monographs,* Provincetown (MA), v.19, p.409-460, 1937.

ANASTASI, A. *Psychological testing.* 6.ed. New York: Macmillan, 1988.

ANGOFF, W.H. Use of difficulty and discrimination indices for detecting item bias. In: BERK, R.A. (Ed.). *Handbook of methods for detecting test bias.* Baltimore (MD): Johns Hopkins University Press, 1982. p.96-116.

BERNARDIN, H.J., WALTER, C.S. Effects of rater training and diary keeping on psychometric error in ratings. *Journal of Applied* Psychology. Washington. D.C., v.62, n.1, p.64-69,1977.

BINET, A., SIMON. T. Méthodes nouvelles pour le diagnostic du niveau intellectuel des anormaux. *L'Année Psychologique.* Paris, v.1 1, p.191-244, 1905.

BIRNBAUM, A. Some latenttrait models and their use in inferring examinee's ability. In: LORD. F.M.. NOVICK, M.R. (Eds.). *Statistical theories of mental test scores.* Reading (MA): Addison-Wesley. 1968.

BORMAN, W.C. Effects of instructions to avoid halo error on reliability and validity of performance evaluation ratings. *Journal of Applied Psychology.* Washington, D.C., v.60, p.556-560. 1975.

CAMPBELL, D.T.. STANLEY, J. *C. Experimental and quasi-experimental designs for research.* Chicago (IL): Rand McNally. 1963.

CAMPBELL. D.T.. FISKE. D.W. Convergent and discriminant validation by the multitrait-multimethod matrix. In: JACKSON. D.N., MESSICK, S. (Eds.). *Problems in human assessment.* New York: McGraw-Hill. 1967.

CARPENTER, PA. JUST. M.A., SHELL, P. What one intelligence test measures: a theoretical account on the processing in the Raven Progressive Matrices Test. *Psychological Review.* Washington, D.C., v.97, n.3. p.404-431. 1990.

CRONBACH, L.J. Coefficient alpha and the internal structure of tests. *Psychometrika.* v. 16, p.297-334, 1951.

CRONBACH, L. J., MEEHL, P.E. Construct validity in psychological tests. *Psychological Bulletin,* Washington, D.C., v.52, p.281-302, 1955.

DUNNETTE, M.D. (Ed.). *Handbook of industrial and organizational psychology.* New York: Wiley, 1983.

EDGERTON, H.A. A table for computing the phi coefficient. *Journal of Applied Psychology,* Washington, D.C., v.44, p.141-145, 1960.

EDWAR DS, A. L. *The social desirability variable in personality assessment and research.* New York: Dryden, 1957.

--------------- . Social desirability and personality test construction. In: BASS. B.M., BERG, I .A. (Eds.). *Objective approaches to personality assessment.* New York: Van Nostrand, 1959.

EDWARDS, AL, WALKER, J.N. Social desirability and agreement response set. *Journal of Abnormal and Social Psychology,* n.62, p.180-183, 1961.

EDWARDS. AL, DIERS.C.J., WALKER, J.N. Response sets and factor loadings on 61 personality scales. *Journal of Applied Psychology.* Washington, D.C., n.46, p.220-225. 1962.

EDWARDS. A.L., WALSH, J.A. The relationship between the intensity of the social desirability keying of a scale and the correlation of the scale with Edwards' SD scale and the first factor loading of the scale. *Journal of Clinical Psychology.* Brandon (VT), n. 19, p. 200-203, 1963.

ELLIS, B.B. Item response theory: a tool for assessing the equivalence of translated tests. *Bulletin of the International Test Commission,* v.18, n.1-2, p.33-51, 1991.

GESELL. A., AMATRUDA, C.S. *Developmental diagnosis.* 2ed. New York: Hoeber-Harper. 1947.

GILLINSKY. A.S. The influence of the procedure of judging on the halo effect. *American Psychologist.* Washington, D.C., v.2. p.309-310, 1947.

GUILFORD. J.P. *Psychometric methods.* New York: McGraw-Hill. 1954.

--------------- . *Personality.* New York: McGraw-Hill. 1959.

GUILFORD. J.P. FRUCHTER. B. *Fundamental statistics in Psychology and Education.* 5,ed. New York: McGraw-Hill, 1973.

GUTTMAN. L. Measurement as structural theory. *Psychometrika.* v.36. n.4. p.329-347, 1971.

HALVERSON. H.M. The acquisition of skill in infancy. *Journal of Genetic Psychology.* Province-town (MA), v.43, p.3-48, 1933.

HAMBLETON. R.K. Adapting tests for use in different cultures: technical issues and methods. *Bulletin of the International Test Commission,* v.18, n.1-2, p.3-32.1991.

HEINEMAN. CE. A forced-choice form of the Taylor Anxiety Scale. Unpublished doctoral dissertation. State University of Iowa. 1952.

HULIN. C.L.. DRASGOW. F.. PARSONS, C.K. *Item response theory,* applications to psychological measurement. Homewood (IL): Dow Jones-Irwin, 1983.

IRONSON, G.H. Use of chi-square and latent trait approaches for detecting item bias. In: BERK, R.A. (Ed.). *Handbook of methods for detecting test bias.* Baltimore (MD): Johns Hopkins University Press. 1982. p.117-160.

JACKSON, D.N. Problems in preparing personality tests and interest inventories for use in multiple cultures. *Bulletin of the International Test Commissions.IH.* n. 1-2. p.94-93, 1991.

JUNG. C.G. *Psychologische Typen.* Zurich: Rascher. 1921,

JURGENSEN, CE. Table for determining phi coefficients. *Psychometrika,* v. 12. p. 17-29, 1947.

KELLEY. T.L. The selection of upper and lower groups for the validation of test items *Journal of Educational Psychology.* Washington, D.C., v.30, p.17-24. 1939.

KNOBLOCK. H., PASAMANICK. B. (Eds.). *Gesell and Amatruda's developmental diagnosis.* 3.ed. New York: Harper and Row. 1974.

KRETSCHMER. E. *Physique and character.* New York: Hartcourt. 1925.

KUDER. G.F.. RICHARDSON. M.W. The theory of estimation of test reliability. *Psychometrika.* n.2. p. 151-160. 1937.

LAURENDEAU. M.. PINARD. A. *Causal thinking in the child:* a genetic and experimental approach. New York: International Universities Press, 1962.

--------------- . *The development of the concept of space in the child.* New York: International Universities Press, 1970.

LORD, F.M. *Applications of item response theory to practical testing problems.* Hillsdale (NJ): Erlbaum. 1980.

MESSICK. S., JACKSON. D.N. Acquiescence and the factorial interpretation of the MMPI. *Psychological Bulletin.* Washington. D.C. v.58. p.299-304. 1961.

MULHOLLAND. T.M.. PELLEGRINO. J.W., GLASER, R. Components of geometric analogy solution. *Cognitive Psychology, v.* 12..p.252-284. 1980.

MURRAY. H.A., et al. *Explorations in personality.* New York: Oxford University Press. 1938.

NEWELL. A.. SIMON. H.A. *Human problem solving.* Englewood Cliffs (NJ): Prentice-Hall. 1972.

NOVICK. M.R.. LEWIS. C. Coefficient alpha and the reliability of composite measurements. *Psychometrika. v.12.* p. 1-13, 1967.

OOSTERHOF, A.O Similarity of various item discrimination indices. *Journal of Educational Measurement.* Washington. D.C. v.13. p.145-150, 1976.

PELLEGRINO. J.W.. MUMAW. R.J.. SHUTE. V.J. Analysis of spatial aptitude and expertise. In: EMBRETSON, SE. (Ed.). *Test design:* developments in psychology and psycho-metrics. 1985. p.45-76.

PIAGET. J. *The origins of intelligence in children.* New York: International Universities Press. 1952.

PINARD. A.. LAURENDEAU, M. A scale of mental development based on the theory of Piaget: description of a project. *Journal of Research in Science Teaching.* New York, v.2. p.253-260, 1964.

POPPER. K.R., ECCLES, J.C. *O eu e seu cérebro.* Brasília: Ed. UnB, 1977.

RASCH. G. An individualistic approach to item analysis. In: LAZARSFELD, P.F., HENRY, N.W. (Eds). *Readings in mathematical social sciences.* Cambridge (MA): MIT Press, p.89-107.

RULON. P.J. A simplified procedure for determining the reliability of a test of split-halves. *Harvard Educational Review.* Cambridge (MA), v.9, p.99-103, 1939.

SHARON, A.T.. BARTLETT. C.J. Effect of instructional conditions in producing leniency on two types of rating scales. *Personnel Psychology.* Durham (NC), v.22, p.251-263, 1969.

SHELDON. W.H. *The varieties of human physique:* an introduction to constitutional psychology. New York: Harper, 1940.

------------------ . *The varieties of temperament:* a psychology of constitutional differences. New **York:** Harper. 1942.

SPEARMAN. C. "General intelligence" objectively determined and measured. *American Journal of Psychology,* Champaign (IL), v. 15, p.201-293. 1904.

STERNBERG, R.J. *Intelligence, information processing, and analogical reasoning:* the com-ponential analysis of human abilities. Hillsdale (NJ): Erlbaum, 1977.

----------------. The nature of mental abilities. *American Psychologist,* Washington. D.C.. v.U. p.214-230. 1979.

----------------. Representation and process in linear syllogistic reasoning. *Journal of Experimental Psychology: General.* Washington. D.C. v.109. p.119-159, 1980.

TAYLOR. E.K.. HASTMAN. R. Relation of format and administration to the characteristics of graphic rating scales. *Personnel Psychology.* Durham (NC), v.9, p.181-206, 1956.

TERMAN. L.M.. MERRILL, M.A. *Stanford-Binet Intelligence Scale:* manual for the third revision. Form L-M. Boston (MA): Houghton Mifflin. I960.

THORNDIKE. E.L. A constant error in psychological ratings. *Journal of Applied Psychology*. Washington. D.C.v.4. p.25-29. 1920.

CAPITULO 4

MEDIDAS ESCALARES

Luiz Pasquali
Instituto de Psicologia
Universidade de Brasília

As medidas escalares em Psicologia se situam dentro do enfoque epistemologico defendido pela psicometria, trabalhando com o modelo da estrutura latente. Elas fazem uso dos conceitos de traço latente, sistema, propriedade e magnitude (vide cap. 3).

Os parâmetros da medida (validade, fidedignidade) também continuam sendo os parâmetros fundamentais na medida escalar. Assim, o conhecimento dos capítulos sobre a teoria da medida e da medida psicométrica se faz necessário para a compreensão do que será dito sobre as medidas escalares.

A medida escalar constitui uma das várias formas que a medida psicométrica pode assumir. Nesta se incluem os testes psicológicos, os inventários, as escalas, etc. As medidas escalares são mais utilizadas na Psicologia Social, especificamente no estudo das atitudes, e também no campo da personalidade, com o intuito de medir traços de personalidade (como, por exemplo, o inventário de Comrey: Escalas de Personalidade de Comrey). Elas se distinguem dos testes e inventários, porque aqueles são de uso mais corrente na avaliação das aptidões (onde há respostas certas e erradas) e estes, no campo da personalidade e da psicopatologia. Além disso, os testes e os inventários, em confronto com as escalas, se apresentam como medidas para as quais existem normas de interpretação, ao passo que para as escalas comumente não são elaboradas tais normas. Na verdade, diferenças essenciais entre estes vários tipos de medidas psicométricas não existem. Há mesmo dúvidas quanto a existirem diferenças importantes entre escalas psicométricas e escalas psicofísicas. A distinção, neste último caso, talvez ainda faça sentido. A escala psicofisica visa escalonar estímulos físicos (através de medida fundamental) que corresponderiam ou produziriam uma escala intervalar psicológica (escala de resposta), sendo as duas relacionadas por alguma lei psicofisica. A escala psicométrica visa escalonar estímulos que expressam um construto psicológico, e seria mais neste sentido restrito

que se usaria mais comumente o conceito de escala. Mas estas distinções se tornam muito tênues, porque, afinal, sempre se escalonam estímulos (itens) observáveis.

Também, a expressão escala é utilizada de múltiplas formas: para designar o nível métrico da medida (escala ordinal, intervalar, etc); para designar um contínuo de números (escala numérica de 5 pontos, por exemplo); para designar os próprios itens de um instrumento, como no caso do diferencial semântico, onde cada item é chamado de escala; para designar diferentes técnicas de construção e uso de instrumentos psicológicos de medida de atitudes (como escala tipo Thurstone, tipo Likert, etc). Todos estes são usos legítimos da palavra escala e, mesmo, não há contradições em tais usos. Embora eles possam trazer algumas dificuldades, normalmente não produzem ambigüidades no tipo de escala que se está falando. O termo, na verdade, originalmente se refere ao fato de que, ao se proceder a uma medida de um atributo empírico, surge uma série de números ordenados à qual é dado o nome de escala numérica. Assim, qualquer medida resultaria numa escala. No caso presente, entretanto, escala é utilizada como uma forma ou técnica de se fazer a medida, especialmente na área das atitudes, como se verá a seguir.

ESCALAS PSICOFÍSICAS

Estas escalas visam verificar e descrever a correlação que existe entre estímulos físicos (som, peso, tamanho, etc.) e a resposta do sujeito. Mais especificamente, qual é o mínimo valor do estímulo que é capaz de produzir uma resposta no organismo (limiar absoluto) e qual é o mínimo de acréscimo no estímulo necessário para produzir no organismo uma resposta diferente da anterior (limiar diferencial). A determinação do limiar absoluto se faz em termos de 50% de percepção de um dado estímulo: o nível de estímulo que é percebido em 50% das vezes é considerado o limiar absoluto ou nível 0 (zero-inicial) da escala de resposta. Para a determinação dos limiares diferenciais, várias leis foram apresentadas na história da Psicologia. Weber (Stevens, 1951) concebeu alei da constante: para produzir uma resposta diferente da anterior, o estímulo deve ser aumentado por uma constante (k) que deve ser determinada empiricamente para cada modalidade de estímulo (peso, som, etc). Como logo se percebeu que esta lei não correspondia muito à observação dos fatos, Fechner (Stevens, 1951) apresentou uma lei logarítmica, na qual a resposta depende de uma constante, diferente para cada modalidade de estímulo, a qual multiplica o logaritmo do estímulo; isto é, para produzir uma resposta diferente da anterior, o estímulo tem que aumentar logaritmicamente: a resposta aumenta aritmeticamente e o estímulo, geometricamente. Outras leis vieram substituir a de Fechner (Stevens, 1951; Guilford,

1951; Guilford, 1954). Stevens, de fato, demonstrou que alguns pressupostos de Fechner não podiam ser mantidos e introduziu novos procedimentos que vieram a se caracterizar como a lei da potência (vide capítulo 2).

Uma exposição detalhada das medidas psicofísicas vai além da intenção deste capítulo. Para tal informação devem ser consultados os trabalhos de Stevens (1951) e Guilford (1954), por exemplo.

Dar-se-ão mais detalhes na exposição das escalas propriamente psicométricas, a seguir. Na apresentação dos vários tipos de escalas, três niveis de preocupação devem ser levados em conta: os procedimentos teóricos, os procedimentos empíricos (experimentais) e os procedimentos analíticos, os quais discriminariam diferentes tipos de escalas psicológicas. Um manual prático para trabalhar com escalas psicométricas é o livro de A.L. Edwards (1957), *Techniques of Attitude Scale Construction.*

O ENFOQUE DE THURSTONE

Caracterizando o **pólo teórico** de sua posição, Thurstone (1927) introduziu o conceito de contínuo psicológico em oposição ao contínuo físico da psicofisica. A diferença é a seguinte: suponha 10 objetos de igual tamanho, mas com pesos diferentes. Estes objetos podem ser ordenados pelo peso de duas maneiras. Primeiro, pode-se usar uma balança e ordenar os objetos pelo seu peso real, produzindo um contínuo físico (através de medida fundamental); mas, segundo, pode-se também pedir a indivíduos, na falta de uma balança, para ordenar os objetos do mais leve ao mais pesado, e esta ordenação constitui um contínuo psicológico de pesos. Esta ordenação psicológica pode ser feita pelos sujeitos comparando os 10 objetos dois a dois, até se chegar à ordem final. Com base neste raciocínio, Thurstone desenvolveu a lei do julgamento comparativo, que poderia ser considerada como introduzindo o conceito de métodos de escalagem psicológica (ou métodos psicométricos em sentido estrito), em oposição aos métodos psicofísicos.

A lei do julgamento comparativo se explicita assim: ao comparar dois estímulos 'i' e 'j' para decidir qual deles é maior (ou 'mais do que' em algum atributo dado, como peso, por exemplo), o sujeito tem que fazer três julgamentos. Primeiro ele tem que avaliar o estímulo '¡', depois, o estímulo 'j' e, finalmente, a diferença i > j. Ao avaliar os dois estímulos individualmente, o sujeito produz um processo discriminante *(discriminalprocess)* e ao avaliar a diferença entre os dois estímulos, ele produz uma diferença discriminante *(discriminal difference).* Contudo, ao fazer esses julgamentos em ocasiões diferentes, o mesmo sujeito não produz o mesmo processo discriminante, de sorte que, daí, resulta uma variabilidade chamada disper-

são discriminante em torno de um processo discriminante modal que corresponde à média dos vários processos discriminantes com referência ao estímulo. Assim, para cada estímulo, sobre o qual existe uma série de julgamentos (muitos sujeitos avaliando o mesmo estímulo ou o mesmo sujeito avaliando o estímulo em muitas ocasiões diferentes: **procedimentos experimentais),** temos um processo discriminante modal e uma dispersão discriminante, isto é, a média e o desvio padrão, dado que os processos discriminantes se distribuem normalmente (suposição razoável).

Ao se fazerem os julgamentos comparativos entre estímulos para definir qual deles é 'maior que' em algum atributo, produz-se uma tabela de freqüências do tipo f, = i > j, conforme Tabela 4-1.

Tabela 4-1 — Freqüência de vezes em que **i** é considerado maior que j por 100 sujeitos (Matriz F_{ij})

	ITENS (i →)	1	2	3	4	5	6
j	1	50	60	45	70	80	95
	2	40	50	30	60	50	80
	3	55	70	50	70	80	90
	4	30	40	30	50	65	85
	5	20	50	20	35	50	60
	6	05	20	10	15	40	50

Obs.: Na diagonal estão os N/2.

Esta matriz F_{ij} pode ser transformada na matriz P,, onde $p_{ij} = f_{ij}/N$, que pode ser transformada na matriz Z_{ij} através da tabela da curva normal, a qual dá as posições escalares dos itens no contínuo do construto, conforme exemplificado nas Tabelas 4-2 e 4-3.

Tabela 4-2 — Matriz P,

	ITENS (i →)	1	2	3	4	5	(.
j	1	.500	.600	.450	.700	.800	.950
	2	.400	.500	.300	.600	.500	.800
	3	.550	.700	.500	.700	.800	.900
	4	.300	.400	.300	.500	.650	.850
	5	.200	.500	.200	.350	.500	.600
	6	.050	.200	.100	.150	.400	.500
	Soma	2.000	2.900	1.850	3.000	3.650	4.600

Tabela **4-3 — Matriz** Z_{ij}

i

	ITENS	1	2	3	4	5	6
j	1	.00 -	.25	-.12 -	.52 .25	.84	1.65
	2	.25	.00	.52 .00	.52 .00	.00	.84
	3	.13	.52	-.52 -	-.39 -	.84	1.28
	4	-.52	-.25	.84 -	1.04	.39	1.04
	5	-.84	.00	1.28		.00 -	.25
	6	-1.65	-.84			.25	.00
	Soma (S)	-3.13	-.32	-3.28 0	-.14	1.82	5.06
	S + 3.28	.15	2.96	0	3.14	5.10	8.34
	Valor =	0	3		3	5	8

Assim, a separação entre os processos discriminantes modais de dois estímulos em termos de desvios normais é

$$E_i - E_j = z_{ij} DP_{ij}$$

Mas, como $DP_{ij} = \sqrt{DP_i^2 + DP_j^2 - 2r_{ij}DP_iDP_j}$ segue que

$$E_i - E_j = z_{ij}\sqrt{DP_i^2 + DP_j^2 - 2r_{ij}DP_iDP_j}\text{ , ou então}$$

$$z_{(ij)} = (E_i - E_j)/\sqrt{DP_i^2 + DP_j^2 - 2r_{(ij)}DP_iDP_j}\text{ ,}$$

onde Z_{ij} = desvio normal, E_i = processo discriminante modal do estímulo i, E_j= processo discriminante modal do estímulo j, DP_i = dispersão discriminante do estímulo i, DP_j = dispersão discriminante do estímulo j e r_{ij} = correlação entre E_i e E_j.

Esta é a fòrmula que determina os **procedimentos analíticos** e permite estabelecer as posições escalares dos estímulos e suas distâncias. Entretanto, para tornar esta fórmula solucionável, Thurstone fez outras suposições, dado que ela tem incógnitas demais; de fato, o único valor calculável a partir dos dados empíricos é o z_{ij}. Das várias suposições que Thurstone fez, como a igualdade das dispersões discriminantes **e a** inexistência de correlação entre as respostas aos dois estímulos, surgiram os famosos cinco casos da lei de Thurstone. O Caso V, por exemplo, é

$$z_{ij} = E_i - E_j,$$

que foi utilizado nos cálculos das tabelas acima para a obtenção dos valores escalares (processos discriminantes modais) dos estímulos. Vale ressaltar que Thurstone oferece testes estatísticos para a verificação da

consistência interna da escala resultante, bem como a verificação da adequação das suposições feitas em cada caso utilizado.

Tendo-se obtido os valores escalares, em termos de desvios padrões, de uma grande série de estímulos, pode-se construir uma escala intervalar, selecionando aqueles (cerca de 20) que se situam a distâncias iguais entre si.

Estes estímulos assim escalonados constituem a escala para a medida das atitudes. Os **procedimentos experimentais** para aferir as atitudes do sujeito consistem em pedir ao mesmo que escolha o item (estímulo) com o qual ele mais concorda, sendo o valor escalar deste item a medida da atitude do sujeito. Ou pede-se para o sujeito escolher os três ¡tens com os quais mais concorda, e a medida da sua atitude será a média dos valores escalares destes três itens.

A construção de escalas a partir desta lei de Thurstone é extremamente laboriosa. Na verdade, ela se torna quase impossível com um número elevado de itens, dado que a comparação dos mesmos 2 **a** 2 aumenta geometricamente o número de comparações a serem feitas. Para 10 estímulos temos (10 x 9)/2 = 45 comparações, e para 100 itens temos (100 x 99)/2 = 4.950. Por isso, Thurstone desenvolveu outras técnicas de construção de escalas de atitude. Uma delas é o método dos intervalos aparentemente iguais (Thurstone, Chave, 1929).

No caso deste método, as afirmações (cerca de 100) sobre um objeto de interesse são impressas em cartões que os sujeitos devem distribuir em 11 pilhas segundo o grau de favorabilidade que, na sua opinião, a afirmação apresenta em relação ao objeto psicológico. As 11 pilhas são erigidas sobre um contínuo de cartões etiquetados de A a K, onde A está ancorado com a expressão 'desfavorável', o K com 'favorável' e o F (o cartão a meio caminho de A e K) com 'neutro'.

O valor escalar dos itens se faz através do cálculo da mediana, tendo como coeficiente de variabilidade o intervalo semi-interquartílico, como na Tabela 4-4.

Tabela 4-4 — Cálculo do valor escalar pelo método dos intervalos aparentemente iguais

AFIRMAÇÕES		CATEGORIAS											ESCALA	Q
		A	B	C	D	E	F	G	H	1	J	K		
		1	2	3	4	5	6	7	8	9	10	11		
1	f	2	2	6	2	6	62	64	26	18	8	4	6.8	1.7
	P	.01	.01	.03	.01	.03	.31	.32	.13	.09	.04	.02		
	pa	.01	.02	.05	.06	.09	.40	.72	.85	.94	.98	1.00		
2	f	0	0	0	10	40	28	50	26	28	24	4	6.9	2.8
	P	.00	.00	.00	.05	.20	.14	.25	.13	.14	.07	.02		
	pa	.00	.00	.00	.05	.25	.39	.64	.77	.91	.98	1.00		

f - freqüência; p - proporção; pa - proporção acumulada

Diversas variantes deste método foram propostas (Ballin, Farnsworth, 1941; Seashore, Hevner, 1933; Edwards, Kilpatrick, 1948; Webb, 1951). O próprio Thurstone (Saffir, 1937) apresentou uma variante que chamava de método dos intervalos sucessivos.

O ENFOQUE DE LIKERT

A técnica de Rensis Likert (1932) talvez seja a mais utilizada na construção de escalas psicométricas e é conhecida, desde que Bird (1940) assim a chamou, como método dos pontos somados *(methodof summated ratings).*

Em seu **pólo teórico,** Likert sustenta que uma atitude (propriedade psicológica) constitui uma disposição para a ação. Esta concepção apresentava dificuldades para Likert na época, dado o enfoque do behaviorismo positivista que defendia a atitude como sendo um simples substituto verbal para a ação concreta. Ele defendeu a atitude como um elemento da personalidade, talvez concebido como um construto hipotético, ao afirmar "se de fato tais elementos existem" — *if, in fact, any such elements exist* (Likert, 1932, p.8). Defendeu igualmente que há uma série de tais construios de personalidade e não um único; novamente uma diatribe espelhando as disputas da época entre unifatoristas e multifatoristas. Likert nem se pôs a questão da magnitude das propriedades psicológicas (atitude, mais especificamente), pois era para ele uma questão já decidida, isto é, as propriedades psicológicas têm magnitudes, por isso é que podem ser medidas.

A preocupação da escala Likert não consiste em procurar determinar o valor escalar dos itens, como pretendia Thurstone, mas verificar o nível de concordância do sujeito com uma série de afirmações que expressem algo de favorável ou desfavorável em relação a um objeto psicológico. As afirmações são respondidas numa escala de 3 ou mais pontos, isto é, o sujeito tem que dizer se concorda, está em dúvida ou discorda com o que a frase afirma sobre o objeto psicológico. O número de pontos na escala de resposta varia de 3 a mais de 10, sendo as mais utilizadas as escalas de 5 e 7 pontos. Aliás, o número de pontos utilizados nas escalas Likert parece ser algo irrelevante. Na pesquisa de Matell e Jacoby (1972), foram utilizadas escalas com 2 até 19 pontos; com exceção das escalas de 2 e 3 pontos (por oferecerem poucos graus de liberdade), em todas as outras a porcentagem de uso dos pontos e o tempo de resposta não foram afetados de modo significativo. Outros estudos já haviam descoberto que o número de pontos da escala, bem como a existência ou não de um ponto neutro, não afeta a consistência interna da escala Likert (Bendig, 1954; Komorita, 1963; Matell, Jacoby, 1971), nem a estabilidade teste-reteste (Jones, 1968; Van der Veer, Howard e Austria,

1970; Goldsamt, 1971; Matell, Jacoby, 1971) e nem a validade concorrente e preditiva (Matell, Jacoby, 1971, 1972).

Os **procedimentos empíricos** consistem em 1) criar um número n de itens sobre um construto psicológico e 2) ter estes itens respondidos por N sujeitos numa escala de n pontos. Para a construção dos itens, vide critérios descritos na parte final deste capítulo.

Os **procedimentos analíticos** visam determinar a seleção final dos itens e a avaliação dos parâmetros psicométricos da escala.

Likert sugere selecionar os itens em termos do seu poder de discriminação de grupos-critério, formados estes à base do escore total que os sujeitos obtêm na escala. Assim, um teste ¥ entre as médias de cada item, obtidas pelos grupos superior e inferior (os 30% escores superiores e 30% inferiores na escala) definem a discriminabilidade dos itens. Entretanto, as análises mais modernas da IRT parecem mais promissoras neste particular, pois elas oferecem até três parâmetros para os itens: discriminação, dificuldade e resposta ao acaso.

Na análise da própria escala, importa verificar a validade e a precisão. Uma análise importante da escala consiste em verificar a unidimensional idade supostada mesma. Tipicamente se utiliza, para tanto, a análise fatorial. A análise da fidedignidade é comumente feita através da análise da consistência interna dos itens através do coeficiente alfa de Cronbach. Mas qualquer das técnicas de validade e precisão podem ser aqui utilizadas.

Fica ainda em dúvida se a escala de Likert produz medidas somente ordinais ou se chegam a ser de intervalo. Na verdade, com os dados empíricos coletados com a escala, pode-se avaliar o valor escalar das categorias utilizadas (os pontos) na escala de resposta (Edwards, 1957) e, a partir daí, utilizar estes valores escalares para as categorias. Tal procedimento, contudo, tira a leveza e a facilidade de trabalhar com as escalas tipo Likert. Edwards e Kenny (1946), aliás, verificaram que escalas construídas no estilo Likert (considerando as categorias 1, 2, 3, etc, como intervalos iguais) correlacionam em torno de 0,90 com escalas de intervalos aparentemente iguais de Thurstone. Concluem ainda que, dada a facilidade de construção e utilização, as escalas tipo Likert se apresentam com grande vantagem sobre as de tipo Thurstone.

O ENFOQUE DE GUTTMAN

Guttman apresentou seu escalograma, para avaliar atitudes, numa série de trabalhos (1944, 1945, 1947, 1950).

A parte **teórica** da técnica supõe que a propriedade psicológica possua magnitude e seja unidimensional. Cada item (indicador comporta-

mental) expressa um nível diferente de magnitude, seguindo uma série monotônica crescente (pelo menos de ordem). De sorte que o conjunto de itens da escala expressa o contínuo da propriedade e que, sendo cumulativos, a aceitação de um item de maior nível implica a aceitação de todos os itens inferiores, isto é, de menor posto.

Os **procedimentos empíricos** da técnica consistem em construir uma série de itens sobre um construto de tal forma que os itens possam ser escalonados cumulativamente. Assim, o sujeito que concorda com um item que expressa um certo nível de atitude com respeito ao construto concordará com todos os itens que têm um nível menor. Desta forma, uma série de itens (cerca de dez) podem ser escalonados do mais fraco ou brando até o mais extremo, produzindo uma escala, pelo menos, ordinal. Sendo isto verdade, basta saber o item mais extremo com o qual o sujeito concorda para podermos reproduzir perfeitamente suas respostas nos outros itens. Na realidade, porém, as coisas não acontecem tão certas assim, de sorte que a reprodução das respostas do sujeito nunca será perfeita. Então se pergunta: qual é o mínimo de reprodutibilidade das respostas aceitável para se poder dizer que uma escala satisfaz o critério de cumulatividade? Com este intuito, Guttman desenvolveu **procedimentos analíticos** para determinar um índice de reprodutibilidade, o qual resulta da comparação entre as respostas correta e incorretamente endossadas. Suponha o seguinte: quatro afirmações sobre um construto psicológico (tendo valor 1 a afirmação mais extrema de atitude) respondidas por seis sujeitos em termos de estar de acordo (valor 1) ou não-acordo (valor 0). A Tabela 4-5 recolhe os resultados fictícios obtidos.

Tabela 4-5 — Dados fictícios para a escala de Guttman

SUJEITOS	AFIRM AÇÕES				SOMA	ERROS
	1	2	**3**	4		
1	1	1 1	0 1	1 1	**3**	1
2 **3**	0 0	0 0	1 1	1	**3**	0 0
4 5	0 0	0	0	0 1	2 1	1
					1	0

Esta tabela é montada de tal forma que nas colunas estão dispostos os itens em ordem decrescente, do mais extremado ao mais brando, em termos de atitude em relação ao construto, e nas linhas estão dispostos os sujeitos, também em ordem decrescente do escore total obtido nos itens (para cada item com o qual está de acordo, o sujeito recebe um ponto). Um item que recebeu o acordo pelo sujeito obtém valor 1 e obtém 0 se o sujeito não o marcou. Assim se forma uma tabela triangular, de tal sorte que

acima da diagonal deveriam aparecer somente 1 e abaixo somente 0. No caso em pauta, como o item #1 é o que expressa a atitude mais extrema em relação ao construto, o sujeito que está de acordo com este item deveria necessariamente marcar todos os outros itens, fato que não ocorreu com o sujeito 1, que marcou o item #1 mas não marcou o item #3. Situações desta natureza provocam a ocorrência de 0 acima da diagonal, o que é contado como um erro. Para o cálculo do índice de reprodutibilidade contam-se todos os erros, isto é, os 0 acima da diagonal, que, no caso, são dois. Assim, o número de valores apropriados na tabela é 20 - 2 = 18. O coeficiente de reprodutibilidade será 18/20 = 0,90. Guttman afirma que o coeficiente deve ser pelo menos de 0,90 para que a escala possa ser considerada adequada.

Escalas tipo escalograma têm a vantagem de avaliar a unidimensionalidade, bem como de apresentar uma garantia de que a amplitude do contínuo do construto está sendo coberta, particularmente se o número de itens for grande (pelo menos 10, sugere Guttman). Contudo, é relativamente raro de se encontrar na literatura esta forma de escalas. Por quê? Primeiramente, nem todos os construtos psicológicos permitem escalabílidade. Por exemplo, em atitude política, o sujeito pode estar de acordo com uma posição do candidato e não com outra, o que poderia inclusive estar indicando falta de unidimensionalidade da escala. Além disso, Guttman não oferece dicas de como elaborar os itens, isto é, o escalograma tem pouco poder heurístico; ele parece ser mais útil para a análise de uma escala do que para guiar a construção dela.

O ENFOQUE DO DIFERENCIAL SEMÂNTICO

Osgood (Osgood, Suci, 1952; Osgood, Suci, Tannenbaum, 1957) desenvolveu o **pólo teórico** de sua posição e uma técnica para medir o conceito de significado *(meaning).* Este é concebido como um estado cognitivo, entendido como um processo de mediação representativa da realidade. O objetivo do Diferencial Semântico consiste em especificar as condições de estímulo e de resposta sob as quais este processo funciona. O processo de mediação é concebido como algo que se intercala entre o estímulo e a resposta, como segue:

$$E \rightarrow r \rightarrow s \rightarrow R$$

O processo mediativo é constituído pela relação r->s, significando que o E (estímulo externo) detona um processo psicológico interno composto de uma reação interna (r) ao E externo, a qual provoca estimulações internas (s), levando este processo s->r a uma ação externa (R). Este processo cognitivo dá o significado ao E para produzir a ação R. Através de estudos fatoriais, Osgood chegou a definir este processo

cognitivo mediativo corno sendo caracterizado por três grandes fatores: o processo apresenta um aspecto avaliatório (emocional), um de poder e um de atividade. Isto quer dizer que o significado varia em termos destes três fatores.

Como é este processo que determina o matiz da ação do sujeito (a resposta R), seria de importante utilidade poder-se medi-lo. É o que pretende fazer o Diferencial Semântico.

Os **procedimentos empíricos** consistem em elaborar uma série de escalas (itens) que devem constituir uma amostra representativa para a descrição de um conceito. Para ser representativa, esta série deve cobrir as dimensões do significado acima assinaladas (os três fatores). Estas escalas são apresentadas, em forma de adjetivos descritivos bipolares, a uma amostra de sujeitos que as avaliam em relação a um conceito, para ver qual o significado que os sujeitos dão a tal conceito, por exemplo, de pai, nação, etc. O formato das escalas ou adjetivos bipolares é apresentado como segue:

PAI

forte __ :__ : __:__ : __:__ :_ fraco
bom _ :__ : __:__ : __:__ :_ mau
ativo _ :__ : __:__ : __:__ :_ passivo

O sujeito deve marcar um dos sete segmentos da linha que vem ancorada com os adjetivos bipolares. Esta marca indica duas tendências: uma direção para um dos dois pólos e uma intensidade, desde que a marca seja colocada mais próxima ou mais longínqua do pólo escolhido da escala.

Os **procedimentos analíticos** consistem em verificar, pelo menos, dois aspectos: o significado que o conceito tem para o sujeito ou grupo de sujeitos em termos dos três fatores e a distância que diferentes conceitos mantêm entre si para estes mesmos sujeitos.

Para efetuar estas análises, os sete segmentos das escalas (itens) são definidos por uma escala numérica da seguinte forma:

- 3 . - 2 . - 1 . 0 . +1 . +2 . +3

Somando-se os escores de cada item por fator e dividindo pelo número de itens no fator, são obtidos os escores dos sujeitos nos três fatores para o conceito envolvido. Assim, 100 sujeitos avaliaram cinco conceitos em nove escalas (três para cada fator), resultando nos dados da Tabela 4-6 (dados fictícios).

Tabela 4-6 — Matriz de seis conceitos avaliados por 100 sujeitos em nove escalas (dados fictícios)

ESCALAS	CONCEITOS				
	Pai	Herói	Destino	Guerra	Paz
Bom Ruim Doce	3	3	0	-3	3
Amargo Agradável	2	3	0	-3	2
Desagradável Forte	2	2	-1	-3	3
Fraco Grande Pequeno	1	3	3 -2	2	3
Poderoso Impotente Ativo	1	3	3	3	1
Passivo Rápido Vagaroso	2	2	2 -2	3	0
Cortante Embotado	1	2	1	-3	1
	1	3		2	1
	0	0		2	1

Assim, o pai é bom, bastante poderoso e algo ativo, ao passo que guerra é ruim, impotente e passiva, etc.

Além dessa descrição dos conceitos, eles podem ser comparados, verificando a distância semântica entre eles através da fórmula geral de distância:

$$D_{il} = \sqrt{\sum d_{il}^2},$$

onde D_{ij} é a distância linear entre os conceitos i e l e d_{ij} é a diferença algébrica entre as coordenadas para os dois conceitos envolvidos. Por exemplo, a distância semântica entre os conceitos Herói e Destino é: $(3 - -l)^2 + (3 - 0)^2 + (3 - -2)^2 = 50$, sendo a raiz quadrada igual a 7,07, a qual constitui a distância semântica entre os dois conceitos. O cálculo destas distâncias permite colocar os conceitos num espaço n-dimensional, formando uma estrutura espacial semântica, como na Figura 4-1, sabendo-se que as distâncias semânticas entre Pai e Herói é de 2,06 e entre Pai e Destino é de 5,59.

As escalas de tipo diferencial semântico têm-se mostrado bastante fidedignas, com índices de precisão teste-reteste variando entre 0,83 a 0,91 (Osgood, Suci, Tannenbaum, 1957), chegando até a 0,97 (Jenkins, Russell, Suei, 1957). Osgood e colaboradores (1957) apresentam também altos índices de validade concorrente do Diferencial Semântico com as escalas de Thurstone (entre 0,74 e 0,82) e de Guttman (da ordem de 0,79).

Para o leitor brasileiro, há uma exposição clara e prática da técnica de Osgood no livro O *Diferencial Semântico: uma Técnica de Medida nas Ciências Humanas e Sociais,* de CA. Alves Pereira (1986), Editora Ática, São Paulo.
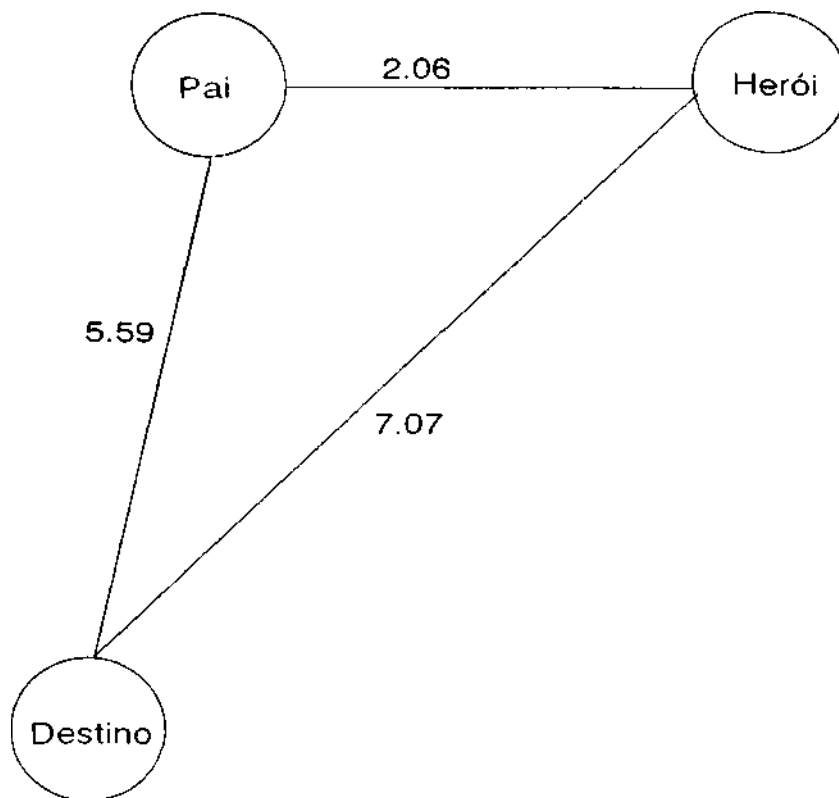
Figura 4-1 — Estrutura espacial semântica de três conceitos.

O ENFOQUE DE FISHBEIN E AJZEN

Fishbein e Ajzen (1975) desenvolvem o seu modelo de escala sob a Teoria da Ação Racional. A teoria e a técnica destes autores estão desenvolvidas no capítulo 14, para o qual remetemos o leitor.

ESCALA MULTIDIMENSIONAL

Pólo teórico

As escalas até aqui apresentadas são ditas unidimensionais, porque elas visam avaliar os sujeitos em apenas um traço psicológico. Falando-se de uma escala multifatorial, neste caso, entender-se-ia um conjunto de várias escalas, cada uma medindo um fator ou traço independentemente.

Entretanto, um objeto psicológico pode ser avaliado sob vários aspectos ou traços simultaneamente. Por exemplo: um candidato à presidência pode ser avaliado em termos de sua filiação partidária (liberal vs. conservador), recebendo uma pontuação nesta escala; ao mesmo tempo e independentemente, ele pode ser avaliado em termos de sua juventude (jovem vs. velho), recebendo nesta escala uma outra avaliação, independente da que recebeu na primeira escala. Teríamos aqui, então, duas escalas unidimen-sionais, produzindo dois escores independentes. Contudo, pode-se pedir uma avaliação simultânea do candidato em termos de ambos os atributos, a saber, filiação partidária e juventude. No primeiro caso, o candidato teria dois escores: um em filiação partidária e outro em juventude. No segundo caso, entretanto, o candidato receberia apenas um escore, mas definido em termos de duas dimensões, que seria ilustrado num espaço bidimensional, onde um ponto é expresso por duas coordenadas. Assim, o escore dele, neste caso, seria expresso como X_{ij} , e não por X_i e X_j. Continuando nesta ilustração, o mesmo candidato poderia ser avaliado numa série de n traços simultaneamente, de sorte que o escore dele poderia ser expresso num espaço n-dimensional, com tantos subscritos quantos os traços sob os quais ele foi avaliado. Estas são as escalas multidimensionais. No caso de uma avaliação em termos de duas dimensões, o escore do candidato poderia cair em qualquer um dos quatro quadrantes que resultam do espaço bidimensional, como na Figura 4-2, onde o candidato $X_i$se situa em (-1,2).
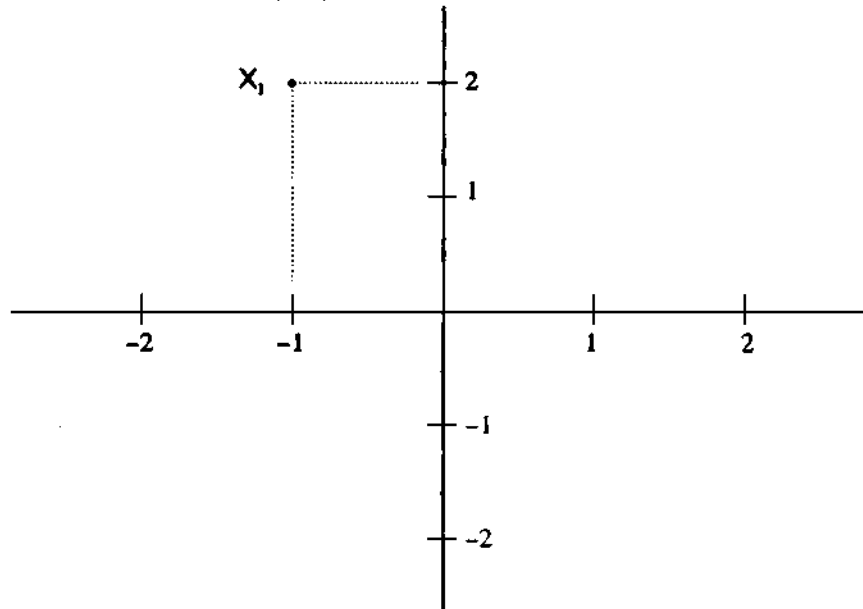
Figura 4-2 — Sujeito X_1 situado em um espaço bidimensional.

Ao ser avaliada uma série de candidatos, cada um deles seria expresso por um ponto que se situaria num dos quadrantes. Os pontos mantêm uma relação de proximidade (distância) entre si expressa simultaneamente com respeito a dois traços: filiação partidária (eixo horizontal) e juventude (eixo vertical). Para n dimensões, o ponto de cada candidato teria, obviamente, proximidades entre si com respeito a n eixos (num espaço n-dimensional).

Pólo experimental

A técnica para levantar os dados de escalas multidimensionais consiste em pedir ao(s) sujeito(s) para avaliar(em) um objeto psicológico (candidato), não em um traço de cada vez, mas em comparar vários objetos psicológicos em vários traços. Por exemplo: Dados os candidatos A B C D, avaliar se os candidatos A e B são mais semelhantes (próximos, iguais, etc.) entre si que os candidatos C e D. Assim, a técnica para a coleta da informação usa termos que se referem a "distância psicológica" ou "proximidade psicológica". Esta proximidade vem designada sob vários termos, tais como parentesco, dependência, associação, complementaridade, substitutividade, proximidade, distância, interação, etc.

Um exemplo poderá ilustrar os procedimentos da técnica das escalas multidimensionais. Suponha quatro candidatos à presidência (A, B, C ,D). Os respondentes reagem à instrução de emparelhar 2 a 2 os candidatos e dizer qual é o candidato preferido entre os dois (i > j). Deste procedimento podem surgir os seguintes dados:

			i		
	CANDIDATOS	A	B	C	D
j	A	7			
	B	5	8-		
	C	3	6	9	
	D				

O candidato A foi preferido 7 vezes a B, 5 vezes a C e 3 vezes a D, etc. Estes números podem ser considerados como indicando distâncias entre os candidatos e serem expressos numa matriz de distâncias. Neste caso, surge uma matriz simétrica, onde a distância d_{ij} é igual a d_{ij} e tendo o valor 0 na diagonal, como segue (note que nem sempre d, deve ser necessariamente igual a d):

$$\begin{bmatrix} 0 & 7 & 5 & 3 \\ 7 & 0 & 8 & 6 \\ 5 & 8 & 0 & 9 \\ 3 & 6 & 9 & 0 \end{bmatrix}$$

Com referência ao **pólo analítico,** a fórmula para cálculo das distâncias é a fórmula euclidiana normalmente utilizada para distâncias, qual seja:

$$d_{ij} = \sqrt{(x_{i1} - x_{j1})^2 + \ldots (x_{ir} - x_{jr})^2}$$

ou, sumariamente, d_{ij} $\sqrt{\sum_{r=1}^{R} (X_{ir} - X_{jr})^2}$,

onde r corresponde ao número de dimensões (fatores) sob as quais os candidatos foram avaliados — no nosso caso, supostamente duas: filiação partidária e juventude.

Para a determinação da dimensionalidade que subjaz às proximidades encontradas entre os objetos psicológicos avaliados (candidatos, no nosso caso), há vários enfoques estatísticos, tanto paramétricos quanto não-paramétricos: *Coombs' unfolding technique* (Coombs, 1964), o modelo de Tucker e Messick (Tucker, Messick, (963), o modelo de Torgerson (1958), o modelo de Ekman (1963), etc. Para tanto, consultem-se Kruskal e Wish (1991) e Delbeke(l968). Existe, igualmente, uma série de programas de computador para as análises com escalas multidimensionais (Kruskal, Wish, 1991, p.79).

A tecnologia das escalas multidimensionais tem sido usada por psicólogos, sociólogos, antropólogos, economistas e educadores (Uslaner, apud Kruskal, Wish, 1991). Seu uso em Psicologia, no entanto, não tem sido muito extenso, apesar do seu caráter promissor na determinação da dimensionalidade nas preferências psicológicas dos indivíduos. O caráter de complexidade estatística talvez seja uma das razões para o pouco uso que se faz das escalas multidimensionais.

PRINCIPIOS DE ELABORAÇÃO DE ESCALAS

Dado o grande uso que se faz das escalas, parece importante apresentar alguns princípios e as etapas relevantes para a elaboração e validação de tais escalas. A maioria dos princípios valem também para a elaboração de qualquer instrumento psicométrico.

Há, pelo menos, três grandes pólos a serem levados em conta ao se tentar elaborar escalas: os pólos teórico, empírico e analítico.

O **pólo teórico** enfoca a questão da teoria que deve fundamentar qualquer empreendimento científico, no caso, a explicitação da teoria sobre o construto ou objeto psicológico para o qual se quer desenvolver um instrumento de medida e a operacionalização do construto em itens. Este pólo explicita a teoria do traço latente, bem como os tipos de comportamentos que constituem uma representação adequada do mesmo traço.

O **pólo empírico** ou experimental define as etapas e técnicas da aplicação de instrumento piloto e da coleta da informação para proceder à avaliação da qualidade psicométrica da escala.

O **pólo analítico** estabelece os procedimentos de análises estatísticas a serem efetuadas sobre os dados para levar a um instrumento válido, preciso e, se for o caso, normatizado.

A Figura 4-3 mostra o organograma destes procedimentos.
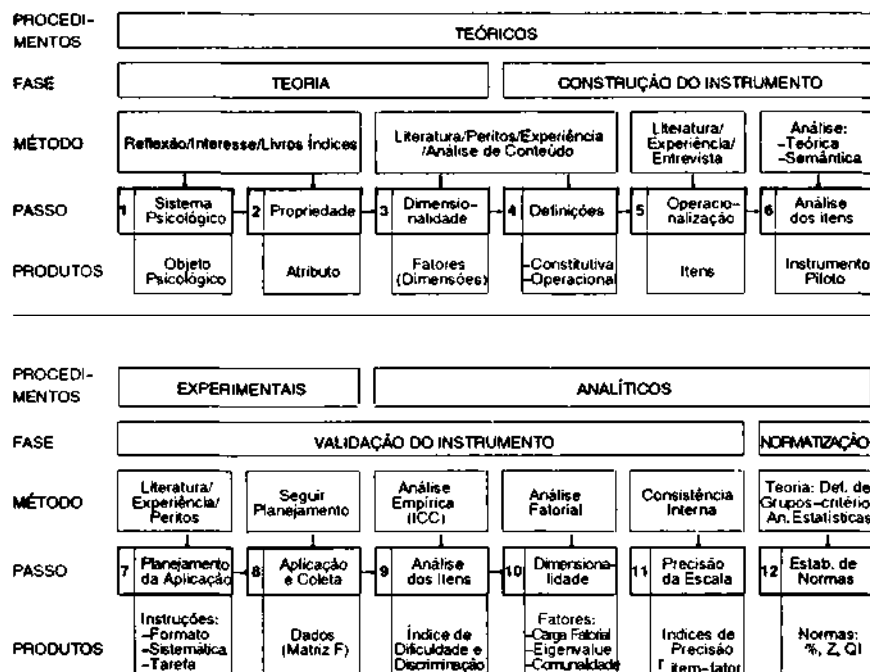
Fig. 4-3 — Organograma para elaboração de escalas de medida psicológica.

Os procedimentos teóricos devem ser elaborados para cada escala, dependendo, portanto, da literatura existente sobre o construto psicológico que a escala pretende medir. A teoria é, infelizmente ainda, a parte mais fraca da pesquisa e do conhecimento psicològico, o que tem como conseqüência a precariedade dos atuais instrumentos psicométricos de medida nesta área. Na verdade, os instrumentos baseados numa teoria psicológica prévia mais elaborada (por exemplo, *Edwards Personal Preference Schedule)* não são dos melhores no mercado. Tal ocorrência explica por que os psicometristas sistematicamente fogem da explicitação de uma teoria preliminar e iniciam a construção do instrumento pela coleta intuitiva e mais ou menos aleatória de uma amostra de itens que dizem possuir *face validity,* isto é, que parecem cobrir o traço para o qual eles querem elaborar o instrumento de medida. Embora isto não pareça muito científico, infelizmente é o que ocorre mais freqüentemente na construção de instrumental psicológico. A inexistência de teorias sólidas sobre um construto não deve ser desculpa para o psicometrista fugir de toda a especulação teórica sobre o mesmo. É obrigação dele levantar, pelo menos, toda a evidência empírica sobre o construto e procurar sistematizá-la e, assim, chegar a uma miniteoria sobre o mesmo para guiá-lo na elaboração de um instrumento de medida para o tal construto. Apesar do avanço e sofisticação estatísticos na psicometria, parece ser esta fraqueza da base teórica que vem maculando a imagem dos procedimentos psicométricos na observação dos fenômenos psicológicos. Na verdade, com uma base teórica coerente e, quanto possível, completa, torna-se viável uma definição dos tipos e características dos comportamentos que irão constituir a representação empírica dos traços latentes e, assim, operacionalizá-los adequadamente (isto é, a construção dos itens se torna coerente e adequada).

Os procedimentos analíticos são encontrados na literatura psicométrica e estatística (vide cap. 6). Acrescentamos aqui apenas uma coletânea de regras úteis para a construção dos itens, que tipicamente constitui a parte mais laboriosa na elaboração das escalas.

A construção dos itens deve seguir certos requisitos, que podem ser explicitados nas regras ou critérios seguintes:

1 — **Critério comportamentali** o item deve expressar um comportamento, não uma abstração ou construto. Segundo Mager (1981), o item deve poder permitir ao sujeito uma ação clara e precisa, de sorte que se possa dizer a ele "vá e faça". Assim, 'reproduzir um texto' é um item comportamental (vá e reproduza...), ao passo que 'compreender um texto' não o é, pois o sujeito não sabe o que fazer com 'vá e compreenda...'.

2 — **Critério de objetividade ou de desejabilidade:** para o caso de escalas de aptidão, os itens devem cobrir comportamentos de fato,

permitindo uma resposta certa ou errada. Para o caso das atitudes e de personalidade em geral, os itens devem cobrir comportamentos desejáveis (atitude) ou característicos (personalidade). O respondente, neste caso, deve poder concordar ou discordar ou opinar sobre se tal comportamento convém ou não para ele, isto é, os itens devem expressar desejabilidade ou preferência.

3 — **Critério da simplicidade:** um item deve expressar uma única idéia. Itens que introduzem explicações de termos ou oferecem razões ou justificativas são normalmente confusos, porque introduzem idéias variadas e confundem o respondente. Por exemplo: "Gosto de feijão porque é saudável". O sujeito pode de fato gostar de feijão, mas não porque seja saudável; assim, ele não saberia como reagir a tal item: se porque o feijão é gostoso ou porque é saudável. O item exprime duas idéias.

4 — **Critério da clareza:** o item deve ser inteligível até para o estrato mais baixo da população meta; daí, utilizar frases curtas, com expressões simples e inequívocas. Frases longas e negativas incorrem facilmente na falta de clareza.

5 — **Critério da relevância** (pertinência, saturação, unidimensionalídade, correspondência): a expressão (frase) deve ser consistente com o traço (atributo, fator, propriedade psicológica) definido e com as outras frases que cobrem o mesmo atributo. Isto é, o item não deve insinuar atributo diferente do definido. O critério diz respeito à saturação que o item tem com o construto, representada pela carga fatorial na análise fatorial e que constitui a covariancia (correlação) entre o item e o fator (traço).

6 — **Critério da precisão:** o item deve possuir uma posição definida no contínuo do atributo e ser distinto dos demais itens que cobrem o mesmo contínuo. Este critério supõe que o item possa ser localizado numa escala de estímulos; em termos de Thurstone, diríamos que o item deve ter uma posição escalar modal definida e um desvio padrão reduzido. Em termos da IRT, este critério representa o parâmetro 'b' (dificuldade) e pode realmente ser avaliado definitivamente somente após coleta de dados empíricos sobre os itens.

7 — **Critério da amplitude:** este critério de fato se refere à escala total e afirma que o conjunto dos itens referentes ao mesmo atributo deve cobrir toda a extensão de magnitude do contínuo deste atributo. Critério novamente satisfeito pela análise da distribuição dos parâmetros 'b' da IRT.

8 — **Critério do equilíbrio:** os itens do mesmo contínuo devem cobrir igualmente ou proporcionalmente todos os segmentos (setores) do contínuo, devendo haver, portanto, itens fáceis, difíceis e médios (para aptidões) ou fracos, moderados e extremos (no caso das atitudes). De fato, os itens devem se distribuir sobre o contínuo numa distribuição que se assemelha à da curva normal: maior parte dos itens de dificuldade

mediana e diminuindo progressivamente em direção às caudas (itens fáceis e itens difíceis em número menor).

9 — **Critério da variedade:** dois aspectos especificam este critério: a) variar a linguagem: uso dos mesmos termos em todos os itens confunde as frases e dificulta diferenciá-las, além de provocar monotonia, cansaço e aborrecimento; b) no caso de escalas de preferências: formular a metade dos itens em termos favoráveis e metade em termos desfavoráveis, para evitar erro da resposta estereotipada à esquerda ou à direita da escala de resposta.

10 — **Critério da modalidade:** formular frases com expressões de reação modal, isto é, não utilizar expressões extremadas, como 'excelente', 'miserável', etc. Assim, ninguém é *infinitamente* inteligente, mas a maioria é *bastante* inteligente.

11 — **Critério da tipicidade:** formar frases com expressões condizentes (típicas, próprias, inerentes) com o atributo. Assim, a beleza não é pesada, nem grossa, nem nojenta.

12 — **Critério da credibilidade** *(face validity):* o item deve ser formulado de modo que não apareça sendo ridículo, despropositado ou infantil. Itens com esta última caracterização fazem o adulto se sentir ofendido, irritado ou coisa similar. Enfim, a formulação do item pode contribuir e contribui (Nevo, 1985; Nevo, Sfez, 1985) para uma atitude desfavorável para com o teste e assim aumentar os erros (vieses) de resposta. Este tema, às vezes, é discutido sob o que se chama de validade aparente *(face validity),* que não tem nada a ver com a validade objetiva do teste, mas pode afetar negativamente a resposta ao teste, ao afetar o indivíduo respondente.

CONCLUSÃO

Apesar dos muitos problemas que ainda existem na teoria da medida em ciências sociais e do comportamento, o uso de escalas, especialmente em Psicologia Social e da Personalidade, além de apresentar uma história de mais de meio século, é ainda muito difundido. Esta ocorrência não pode ser considerada fortuita, mas deve proceder do fato de que as medidas escalares são capazes de produzir conhecimento válido nas ciências do comportamento. As várias técnicas expostas (Likert, Thurstone, Guttman, etc.) têm apresentado razoável consistência, tanto em sua estrutura interna quanto nos resultados obtidos através delas. Todas essas técnicas, na verdade, oferecem procedimentos estatísticos que permitem avaliar essa consistência interna. Quanto à consistência dos resultados que produzem, a situação das escalas existentes e as próprias técnicas propostas para a sua construção não aparecem ainda como

empolgantes. É possível, e quiçá provável, que este fenômeno se deva em grande parte à falta de definição mais precisa destas mesmas técnicas quanto aos procedimentos teóricos envolvidos na elaboração dos instrumentos. Há uma preocupação grande, e louvável, referente à adequação dos procedimentos estatísticos, mas estes não dão dicas fundamentais quanto ao verdadeiro problema da escala, que é a construção de um instrumento válido, isto é, que de fato esteja medindo algo de psicologicamente relevante. Sem uma boa teoria psicológica que a fundamente, a escala pode até aparecer estatisticamente perfeita e consistente, mas medindo nada de relevante ou medindo algo desconhecido.

REFERÊNCIAS BIBLIOGRÁFICAS

ALBANESE. MA., FORSYTH, R.A. The one-, two- and modified two-parameter latent trait models: an empirical study of relative fit. *Educational and Psychological Measurement.* Durham (NC), v.44 n.2, p.229-246, 1984.

BALLIN. M.. FARNSWORTH, P.R. A graphic rating method for determining the scale values of statements in measuring social attitudes. *Journal of Social Psychology.* Provincetown (MA), v.13, p.323-327.1941.

BENDIG. AW. Reliability and the number of rating scale categories. *Journal of Applied Psychology.* Washington, D.C., v.38, p.38-40. 1954.

BIRD, C. *Social psychology.* New York: Appleton-Century-Crofts. 1940.

BIRNBAUN, A. Some latent trait models and their use of inferring an examinee's ability. In: LORD, F.M., NOVICK, M.R. (Eds.). *Statistical theories of mental test scores.* Reading (MA): Addison-Wesley, 1968.

CAMPBELL, N.R. *An account of the principles of measurement and calculations.* London: Longmans Green, 1928.

--------------. Symposium: Measurement and its importance for philosophy. *Proceedings Aristot. Society Suppl..* London, v.17, p.121-142. 1938.

COOMBS, C.H. *A theory of data.* New York: Wiley. 1964.

DE BRUYNE, P., HERMAN, J„ DE SCHOUTHEETE, M. *Dinâmica da pesquisa em ciências sociais.* .Ved. Rio de Janeiro: Francisco Alves. s.d.

DEBETS. P.. BROUWER. E.. SIJTSMA, K., MOLENAAR, I.W. MSP: a computer program for item analysis according to a nonparametric IRT approach. *Educational and Psychological Measurement.* Durham (NC), v.49. n.3. p.609-613, 1989.

DEBETS. P., BROUWER, E. *MSP:* a program for Mokken Scale analysis for polychotomous items. Groningen: IEC ProGAMMA, 1989.

DELBEKE, L. *Construction of preference spaces.* Louvain: Publications of the University of Louvain, 1968.

EDWARDS, AL. *Techniques of attitude scale construction.* New York: Appleton-Century-Crofts. 1957.

EDWARDS. AL., KENNY, K.C. A comparison of the Thurstone and Likert techniques of attitude scale construction. *Journal of Applied Psychology,* Washington, D.C., v.53, p. 72-83, 1946.

EDWARDS, A.L, KILPATRICK, F.P. A technique for the construction of attitude scales. *Journal of Applied Psychology,* Washington, D.C., v.32, p.374-384, 1948.

EKMAN. P. (s.n.t.) 1963.

FISHBEIN, M. AJZEN, I. *Beliefs, attitude and intention:* an introduction to theory and research. Reading (MA): Wesley, 1975.

GOLDSAMT, M.R. Effects of scoring method and rating scale length in extreme response style measurement. College Park (MD): University of Maryland, 1971. (Unpublished doctoral dissertation)

GUILFORD, J.P. *Psychometric methods.* 2ed. New York: McGraw-Hill, 1954.

GUTTMAN, L. A basis for scaling qualitative data. *American Sociological Review,* Washington, D.C..V.9, p.139-150. 1944.

----------------. The basis for scalogram analysis. In: STOUFFER, SA. et al. *Measurement and prediction.* Princeton (NJ): Princeton University Press, 1950. p.60-69.

. The Cornell technique for scale and intensity analysis. *Educational and Psycholog ical Measurement.* Durham (NC), v.7, p.247-280, 1947.

----------------. The problem of attitude and opinion measurement. In: STOUFFER, S.A. et al. *Measurement and prediction.* Princeton (NJ): Princeton University Press. 1950. p.46-59.

---------------. On Festinger's evaluation of scale analysis. *Psychological Bulletin.* Washington, D.C., V.44, p.451-465, 1947.

----------------. *Questions and answers about scale analysis.* Research Branch, Information and Education Division, Army Service Forces. Report D-2. 1945.

JENKINS. J.J., RUSSELL, W.A., SUCI. G.J. An atlas of semantic profiles for 360 words. In: *Studies on the role of language in behavior.* Tech. Rep. No. 15. Minneapolis: University of Minnesota, 1957.

JONES, R.R. Differences in response consistency and subject's preferences for three personality inventory response formats. *Proceedings of the 67th Annual Convention of the American Psychological Association,* v.3, p.247-248, 1968.

KOMORITA. S.S. Attitude content, intensity, and the neutral point on a Likert scale. *Journal of Social Psychology.* Provincetown (MA), v.61, p.327-334, 1963.

KRUSKAL, J.B., WISH, M. *Multidimensional scaling.* Newbury Park (CA): Sage Publications, 1991.

LIKERT, R. A technique for the measurement of attitudes. *Archives of Psychology* . v. 140, 1932.

LORD, F.M. A theory of test scores. *Psychometric Monograph,* Chicago (IL), n.l, 1952.

---------------- .*Applications of item-response theory to practical testing problems.* Hillsdale (NJ): Lawrence Erlbaum, 1980.

MAGER, R.F. *Medindo os objetivos de ensino ou "conseguiu um par adequado".* Porto Alegre: Globo, 1981.

MATELL, M.S., JACOBY, J. Is there an optimal number of alternatives for Likert-scale items? *Journal of Applied Psychology.* Washington, D.C., v.56, n.6, p.506-509, 1972.

----------------. Is there an optimal number of Likert scale items? Study I: Reliability and validity. *Educational and Psychological Measurement,* Durhan (NC), v.31, p.657-674, 1971.

MOLENAAR, I.W., SIJTSMA, K Mokken's approach to reliability estimation extended to multicategory items. *Kwant'rtatieve Methoden.* Rotterdam, v.9 n.28, p.l 15-126,1988.

NEVO, B. Face validity revisited. *Journal of Educational Measurement,* Washington. DC, v.22, p.287-293, 1985.

NEVO, B., SFEZ, J. Examinees' feedback questionnaires. *Assessment and Evaluation in Higher Education,* Bath, v. 10, p.236-249, 1985.

OSGOOD, C.E., SUCI, G.J., TANNENBAUM, PH. *The measurement of meaning.* Urbana (IL): University of Illinois Press, 1957.

----------------. A measure of relation determined by both mean difference and profile information. *Psychological Bulletin,* Washington, D.C., v.49, p.251-262.

PEREIRA, CA. Alves. *O diferencial semântico:* uma técnica de medida nas ciencias humanas e sociais. São Paulo: Ática, 1986.

POPPER, K.R. *A lógica da pesquisa cientifica.* São Paulo: Cultrix, 1972.

RASCH, G. *Probabilistic models for some intelligence and attainment tests.* Copenhagen: Danish Institute for Educational Research, 1960.

SAFFIR, M.A. A comparative study of scales constructed by three psycho-physical methods. *Psychometrika,* v.2, p. 179-198, 1937.

SEASHORE, R.H., HEVNER, K. A time-saving device for the construction of attitude scales. *Journal of Social Psychology,* Provincetown (MA), v.4, p.366-372, 1933.

SIEGEL, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

SIJTSMA, K., DEBETS, P., MOLENAAR, I.W. Mokken scale analysis for polychotomous items: theory, a computer program and an empirical application. *Quality & Quantity,* **apud** DEBETS, P.. BROUWER, E., SIJTSMA, K., MOLENAAR, I.W. MSP: a computer program for item analysis according to a nonparametric IRT approach. *Educational and Psychological Measurement,* Durham (NC), v.49 n.3, p.609-613, 1989.

STEVENS, S.S. Mathematics, measurement, and psychophysics. In: STEVENS, S.S. (Ed.). *Handbook of experimental psychology.* New York: Wiley, 1951. p.1-49.

THURSTONE, L.L. A law of comparative judgment. *Psychological Review,* Washington, D.C., V.34, p.273-286, 1927.

----------------. Psychophysical analysis. *American Journal of Psychology,* Champaign (IL), v.38, p.368-389, 1927.

--------------- . The method of paired comparisons for social values. *Journal of Abnormal and Social Psychology,* v.21, p.384-400, 1927.

--------------- . Equally often noticed differences. *Journal ot Educational Psychology.* Washington, D.C.,v.l8,p.289-293, 1927.

THURSTONE. L.L, CHAVE, E.J. *The measurement of attitude.* Chicago (IL): University of Chicago Press, 1929.

TORGERSON, W.S. *Theory and methods of scaling.* New York: Wiley, 1958.

TUCKER, L.R., MESSICK, S. An individual differences model for multidimensional scaling. *Psychometrika,* v.28, p.333-367, 1963.

VAN DÉR VEER, F., HOWARD, K.I., AUSTRIA, A.M. Stability and equivalence scores based on three different response formats. *Proceedings of the 78th Annual Convention of the American Psychological Association,* v.5, p.99-100, 1970.

WEBB, S.C. A generalized scale for measuring interest in science subjects. *Educational and Psychological Measurement,* Durham (NC), v.ll, p.456-469, 1951.

CAPITULO 5

ESCALAGEM PSICOLÓGICA UNIDIMENSIONAL

Jorge de Souza
Departamento de Economía Universidade
Federal de Pernambuco

Desde os primeiros estudos da psicofisica alemã, na segunda metade do século XIX, os psicometristas lidam com o problema de ordenação de um conjunto de estímulos ou objetos psicológicos relativamente a uma de suas propriedades. Dá-se o caso, por exemplo, quando os sujeitos do experimento, juizes ou examinandos, são instados a manifestarem-se sobre qual o sinal sonoro, entre dois sinais a eles submetidos, é o mais intenso. Postos, assim, diante de um conjunto de sinais sonoros expostos dois a dois, os juizes dão as suas respostas, e o objetivo do experimento é obter uma ordenação ascendente dos estímulos quanto às suas intensidades. Essa mesma ordenação é comparada com a correspondente ordem natural sob controle do pesquisador. Quanto mais elevada for a compatibilidade entre as duas seqüências, maior é a acuidade perceptiva do examinando.

Sob esse enfoque, os diversos sentidos da percepção humana, como a visão, a audição, a sensação de peso, etc, foram investigados na psicofisica e, a partir deles, estabeleceram-se as chamadas **leis da psicofisica.**

Retomando essas idéias no século XX, Louis Thurstone, o grande psicossociometrista norte-americano, concebeu uma situação mais geral e pôde estender as idéias da psicofisica a uma classe mais ampla de problemas que envolviam o conceito de **traço psicológico.** O traço psicológico é uma propriedade ou uma variável de caráter latente — e, portanto, não diretamente observável — ligada a algum sujeito ou objeto de investigação. Enquadra-se nessa ordem, por exemplo, a variável latente que leva os consumidores a manifestarem suas preferências entre diversas marcas de um produto ou os cidadãos a escolherem seus políticos preferidos num elenco de homens públicos.

Algumas das teorias psicométricas sobre o tratamento deste tema são o objeto deste capítulo.

Para o propósito de formalização matemática, considere-se o conjunto $\{O_1, 0_2, ..., O_m\}$ constituido pelos m **objetos psicológicos** ou **estímulos** $O_1, 0_2, ..., O_m$. Esses estímulos, quando submetidos aos julgamentos de n **juizes** ou examinandos J_1 J_2, ..., J_n, suscitam reações comparativas relativamente ao **traço psicológico,** psicofisico ou paranormal que eles representam. Dois modelos estatísticos de reação dos juizes aos estímulos ressaltam neste capítulo:

I) se aos juizes são apresentados os pares de estímulos (O_i, O_j) (¡>j), eles enunciam suas preferências, que podem ser dos dois tipos seguintes:

- O_i]O_j (o juiz prefere O_i a O_j)
- O_i]O_i (o juiz prefere O_i a O_i), não sendo consideradas aqui nem a omissão da opinião e nem tampouco a indiferença no julgamento;

II)se aos juizes são apresentados todos os m estímulos O_1,, O_2, ..., O_m, eles manifestam suas preferências através de uma ordenação decres cente na intensidade com que o traço psicológico, segundo seus juízos, está neles contido. Essa ordenação genérica é representada pela seqüên cia O_{i1}] O_{i2}] ...] O_{im}, onde (i1, ¡2, ..., im) é uma permutação qualquer dos objetos.

O primeiro tipo de reação dá origem ao chamado **modelo dos julgamentos comparativos** e o segundo nomeia o denominado **modelo dos julgamentos categorizados.** Ambos são atribuídos ao engenho estatístico de Louis Thurstone (1927).

Duas situações, bem simples e cotidianas, podem ilustrar o que acima se descreveu:

I) em uma **pesquisa política,** os estímulos podem representar alguns líderes políticos, os juizes são uma amostra de eleitores e o traço psicológico medido pode ser a **preferência ideológica;**

II) Em uma **pesquisa mercadológica,** os estímulos podem signi ficar as diversas marcas de um produto, os juizes constituem uma amostra de consumidores e o traço psicológico pode ser a simples **preferência pela marca.**

As duas situações enunciadas escondem, na realidade, algumas questões relevantes no estudo dos problemas comportamentais através de modelos estatísticos. Com efeito, em ambos os exemplos fica patenteada a idéia, bastante geral, de que se lida com uma **amostra de examinandos** e não com a sua **totalidade** ou a **população.** Esta restrição conduz, é claro, a um problema de inferência estatística concernente à indução

empírica. Quanto aos estímulos, muitas vêzes, também se lida com uma amostra obtida de uma população de infinitas possibilidades. Diz-se, neste caso, que os estímulos escolhidos constituem **manifestações** do traço psicológico estudado e que a população de todas as manifestações possíveis do traço é o seu **universo de conteúdo.** Há, desse modo, um segundo tipo de problema inducional, chamado por Hotelling de **inferência psicológica** (Hotelling, 1933).

O modelo dos julgamentos comparativos admite a existência de um ***continuum* psicológico,** onde todo examinando ou juiz faz a sua avaliação do conteúdo do traço contido em cada estímulo ou objeto psicológico. Esse mecanismo de avaliação é inconsciente, e cada juiz por eles manifesta a sua preferência após comparar as avaliações efetuadas no *continuum* psicológico. As avaliações do conteúdo, independentemente dos juizes que as efetuam, são denominadas de **valores de escala do estímulo.** Dessa maneira, um valor de escala seria a medida verdadeira do conteúdo do traço existente no estímulo. Entretanto, há que se convir que as avaliações do conteúdo do traço em um dado objeto psicológico podem variar de juiz a juiz ou, para um mesmo juiz, flutuar segundo o instante de medição. No primeiro caso, está-se estudando um problema de Psicologia Diferencial e, no segundo, um problema de Psicologia Comportamental.

O modelo dos julgamentos comparativos pode ser imaginado, segundo as descrições anteriores, como se a avaliação de um juiz genérico a um estímulo O_i fosse uma variável aleatória Xi cuja esperança matemática $E[X_i]$ representa o valor de escala do objeto psicológico O_i (i - 1, 2, ..., m). Essa conclusão implica dizer que as avaliações Xi dos juizes flutuam em torno do valor de escala μ_i a menos de um erro aleatòrio e, de esperança matemática nula, ou seja, que

(5.1) $$X_i = \mu_i + \varepsilon_i \; ,$$

onde $E[\varepsilon_i] = 0$ e sendo μ_i desconhecido (i = 1, 2,..., m).

A regra decisoria, de caráter estatístico, além de inconsciente ou latente e que possibilita a um juiz genérico optar por um dos estímulos O_i ou P_j é a seguinte: "O_i] O_j, ou seja, prefere O_i a O_j, se e somente se $X_i > X_j$ onde X_i e X_j são as avaliações do traço psicológico feitas pelo juiz e que, na realidade, são desconhecidas". Essa regra de decisão estabelece, assim, uma relação de correspondência entre **plano psicológico** ou **mental** e o **plano de opiniões,** de acordo com o esquema seguinte:

PLANO PSICOLÓGICO	PLANO DE OPINIÕES
$X_i > X_j$	O_i] O_j
$X_i < X_j$	O_i [O_j

Considerando, agora, a natureza probabilistica da decisão mental a partir das avaliações individuais dos juizes, pode-se estabelecer a seguinte equação para caracterizar o modelo dos julgamentos comparativos:

$$Prob[O_i]O_j] = Prob[X_i > X_j] \qquad (5.2)$$

isto é, são iguais as probabilidades da preferência de O_i relativamente a Oj e do evento[$X_i < X_i$]que expressa o fato de o valor de escala de O_i ser maior do que o correspondente valor de escala de O_j.

O modelo se completa, desde o ponto de vista estatístico, com a especificação de algumas propriedades adicionais relativas às distribuições de probabilidades das m variáveis aleatórias $\varepsilon_1, \varepsilon_2, ..., \varepsilon_m$ representativas dos erros. Louis Thurstone formulou, em 1929, as hipóteses de que os m erros $\varepsilon_1, \varepsilon_2, ..., \varepsilon_m$ têm distribuições normais, mediante as seguintes especificações numéricas:

i) $E[\varepsilon_i] = 0$ (i = 1, 2,..., m), isto é, os erros têm esperanças matemáticas nulas;

ii) $Var[\varepsilon_i] = \sigma_i^2$ (i = 1,2 ..... m), ou seja, os erros ε_i, têm variâncias iguais a $\sigma_i^{2,''}$;

iii) $Corr[\varepsilon_i, \varepsilon_j] = \rho_{ij}$ (i, j = 1, 2,..., m), isto é, os erros ε_i, e ε_j têm coeficientes de correlação igual a ρ_{ij}

Sob tais hipóteses, laborando em torno da equação 5.2, pode-se escrever que

$$Prob[O_i]O_j] = Prob[X_i > X_j] = Prob[X_i - X_j > 0] \qquad (5.3)$$

$$= Prob\left[\frac{(X_i - X_j) - (\mu_i - \mu_j)}{\sqrt{\sigma_i^2 + \sigma_j^2 + 2\rho_{ij}\sigma_i\sigma_j}} > -\frac{(\mu_i - \mu_j)}{\sqrt{\sigma_i^2 + \sigma_j^2 + 2\rho_{ij}\sigma_i\sigma_j}}\right]$$

Tendo em vista que a variável aleatória correspondente ao primeiro membro da desigualdade tem distribuição normal (0,1), resulta que

$$Prob\left[N(0,1) > -\frac{(\mu_i - \mu_j)}{\sqrt{\sigma_i^2 + \sigma_j^2 + 2\rho_{ij}\sigma_i\sigma_j}}\right] = \Pi_{ij} \,,$$

onde Π_{ij} = Prob [Oi > Oj] representa a probabilidade da preferência O_i] O_j

Agora, fazendo

$$Z_{ij} = \frac{-(\mu_i - \mu_j)}{\sqrt{\sigma_i^2 + \sigma_j^2 + 2\rho_{ij}\sigma_i\sigma_j}} \qquad (5.4)$$

, tem-se que

(5.5)

$$Prob\,[N(0,1) > Z_{ij}] = \Pi_{ij} \quad (i,j = 1,\, 2,\, \ldots,\, m)$$
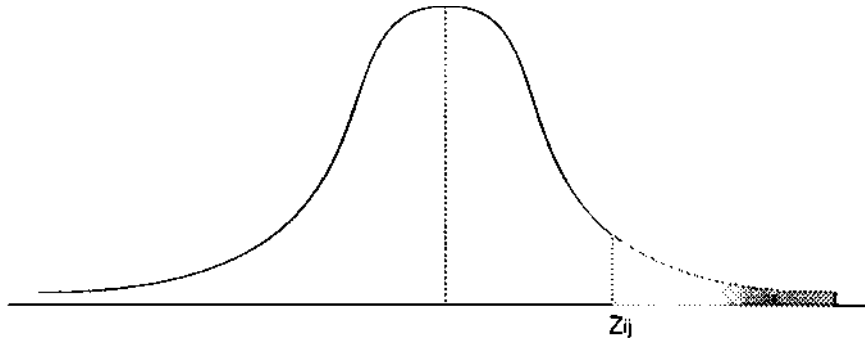
Fig. 5-1

Dessa maneira, a partir da última convenção estabelecida em (5.4), pode-se escrever que

$$\mu_i - \mu_j = Z_{ij}\sqrt{\sigma_i^2 + \sigma_j^2 + 2\rho_{ij}\sigma_i\sigma_j} \quad (j > i)$$

(5.6)

Ao modelo por esse modo caracterizado dá-se o nome de **modelo de julgamentos comparativos** de Thurstone.

O sistema de equações estabelecido em (5.4) é constituído por $\frac{m(m-1)}{2}$ equações e pelas 2m incógnitas $\mu_1 \ldots \mu_m$ e $\sigma_1^2 \ldots \sigma_m^2$, adicionais às $\frac{m(m-1)}{2}$ incógnitas ρ_{ij}, referentes às correlações, ou seja, o sistema é subidentificado por possuir $\frac{m(m-1)}{2}$ equações e $(\frac{m(m-1)}{2} + 2m)$ incógnitas.

O seu caso particular mais simples pressupõe que sejam iguais as variâncias dos erros e nulas as correlações entre eles, isto é, $\sigma_1^2 = \sigma_2^2 = \ldots \sigma_m^2 = \sigma^2$ e $\rho_{ij} = 0 \;\; (i>j)$. Nessas circunstâncias, pode-se escrever o sistema de equações anterior como $\mu_j - \mu_i = z_{ij}\sqrt{2\sigma^2}$ ou, ainda, como $\frac{\mu_j}{\sqrt{2\sigma^2}} - \frac{\mu_i}{\sqrt{2\sigma^2}} = z_{ij}$, donde se conclui que

(5.7) $$\mu_j^* - \mu_i^* = z_{ij} \quad (j > i)$$

O fator $\frac{1}{\sqrt{2\sigma^2}}$ apenas afeta multiplicativamente os valores de escala e pode ser desprezado desde que se interpretem os resultados numa escala ordinal. Dessas digressões resulta que os m valores de escala

devem ser estimados através de um sistema superidentificado formado pelas $\frac{m(m-1)}{2}$ equações

$$\mu_j - \mu_i = z_{ij} \quad (j > i).$$

(5.8)

Ele pode ser resolvido, por exemplo, através do método dos mínimos quadrados, onde deve ser minimizada a função $\psi = \sum[(\mu_i - \mu_i) - Z_{ij}]^2$. Derivando-se relativamente aos valores de escala, encontra-se que

$$\frac{\partial\psi}{\partial\mu_1} = \frac{\partial}{\partial\mu_1}\{[(\mu_2-\mu_1)-Z_{12}]^2 + [(\mu_3-\mu_1)-Z_{13}]^2 + ...\} =$$

$$= -2[(\mu_2-\mu_1)-Z_{12}]-2[(\mu_3-\mu_1)-Z_{13}]... = -2\sum_{j>1}[(\mu_j-\mu_1)-Z_{1j}]$$

$$\frac{\partial\psi}{\partial\mu_2} = \frac{\partial}{\partial\mu_2}\{[(\mu_3-\mu_2)-Z_{23}]^2 + [(\mu_4-\mu_2)-Z_{24}]^2 + ...\} =$$

$$-2[(\mu_3-\mu_2)-Z_{23}]-2[(\mu_4-\mu_2)-Z_{24}]... = -2\sum_{j>2}[(\mu_j-\mu_2)-Z_{2j}]$$

e assim sucessivamente.

Em geral, tem-se que $\frac{\partial\psi}{\partial\mu_i} = -2\sum_{j>i}[(\mu_j-\mu_i)-z_{ij}]$ e, portanto, os estimadores mínimo-quadráticos $\hat{\mu}_1, \hat{\mu}_2, ..., \hat{\mu}_m$ dos valores de escala dos objetos psicológicos serão dados pela solução das equações

$\sum_{i>j}[(\hat{\mu}_i-\hat{\mu}_j)-Z_{ij}]$ = 0 (i=1, 2, ... m-1), ou seja:

$\sum \hat{\mu}_i - \sum \hat{\mu}_j = \sum z_{ij}$ (i=1, 2

.... m-1)

$$\sum_{i>j}\hat{\mu}_i-(m-1)\hat{\mu}_j = \sum_{i>j}z_{ij}.$$

Simétricamente, a soma poderia ser obtida com j < i e isto eqüivaleria a escrever

$$\sum_{j>i}\hat{\mu}_j-(m-1)\hat{\mu}_j = \sum_{j<i}Z_{ij}$$

Adicionando essas duas

$$\sum \hat{\mu}_i-(m-1)\hat{\mu}_j = \sum Z_{ij} \quad (i, j \neq i)$$

equações, tem-se que

ou ainda, somando e subtraindo $\hat{\mu}_i$, do primeiro membro dessa última equação

$$\sum_{j=1}^{m}\hat{\mu}_j-m\hat{\mu}_j = \sum_{i,j\neq i}Z_{ij}$$

Impondo-se a restrição de que $\sum^{m} \hat{\mu}_i = 0$, o que equivale apenas a mudar a origem dos valores de escala, tem-se que $-m\hat{\mu}_j = \sum z_{ij}$.

Convencionando-se, agora, sem nenhuma perda de generalidade, que $z_{ii} = 0$ (¡=1,2, ..., m), tem-se que $\hat{\mu}_i = -\frac{1}{m}\sum^{m} z_{ij}$, ou seja, $\hat{\mu}_i = -\bar{z}_i$ ou, finalmente, em razão do mesmo argumento de simetria sobre os valores de z:

(5.9) $\hat{\mu}_i = -\bar{z}_i$ (/ = 1 , 2 ....m).

Esta última expressão permite concluir que o valor de escala estimado $\hat{\mu}_i$ do objeto psicológico 0¡ (i = 1,2,..., m) só pode ser interpretado numa escala ordinal e seu valor é encontrado considerando as freqüências relativas $\hat{\Pi}_{ij}$ = Freq Rel [O_i] Oj das preferências que sobrepõem O, ao estímulo O, e aos correspondentes z_{ij}, valores da curva normal padronizada, de acordo com a Figura 5-1.0 valor de escala $\hat{\mu}$, é a média

$z_i = -\frac{1}{m}\sum_{j=1}^{m} z_{ij}$, onde se convencionou que $z_{ii} = 0$.

O MODELO DOS JULGAMENTOS CATEGÓRICOS

Thurstone seguiu ainda uma outra ordem de idéias, a fim de explicar o processo decisòrio de preferências dos juizes ou examinadores. Desse modo, se no modelo dos julgamentos comparativos os juizes faziam avaliações das intensidades dos estímulos no chamado *continuum* psicológico, agora, seguindo uma outra formulação, Thurstone admitiu que os juizes julgavam os estímulos localizando-os em um conjunto ordenado de categorias representadas por intervalos do mesmo *continuum* psicológico. Nesse caso, deve-se atribuir valores de escala não só aos estímulos como, igualmente, aos limites que definem as diversas categorias do *continuum* psicológico onde os estímulos são classificados. Suponha, assim, em razão disso, que os m estímulos ou objetos psicológicos O_1, O_2 .... O_m devam ser classificados pelos juizes em um dos seguintes k intervalos $I_1, I_2, ..., I_k$ que constituem uma partição do *continuum* psicológico $(-\infty, c_1], (c_1, c_2], (c_2, c_3], ..., (c_{k-2}, c_{k-1}]$, $(c_k, +\infty)$, naturalmente ordenados de modo ascendente, sendo essa ordenação simbolizada pela seqüência $I_1 < I_2 < ..., < I_k$.

Por outro lado, sendo μ, o valor de escala do estímulo O_i(i = 1, 2, ..., m), a regra decisòria mental de cada juiz é estabelecida pela seguinte equivalência lógica: "O estímulo O_i é classificado em um dos j primeiros

intervalos $(-\infty, c_1], (c_1, c_2], \dots (c_{j-1}, c_j]$ se e somente se seu valor de escala μ_i é menor ou igual ao valor extremo c_j do intervalo de maior intensidade".

Desde um ponto de vista eminentemente simbólico, essa regra pode ser representada do seguinte modo: "$O_i \in I_1 \cup I_2 \cup \dots \cup I_j$ se e somente se $\mu_i \leq c_j$, onde $I_j = (c_{j-1}, c_j]$ é o j-ésimo intervalo da partição".

Impõe-se considerar, ainda, no referente a esse novo modelo, que os k intervalos I_1,, I_2,..., I_k em que são classificados os m estímulos O_1, O_2, ..., O_m, podem ser vistos como representantes de uma ordenação em que, sendo k < m, os estímulos são classificados nos lugares 1º, 2°.,... e k-ésimo, havendo, portanto, a necessidade de alguns empates. No caso em que se tem k = m, ou seja, quando o número de intervalos classificató-rios coincide com o número de estímulos, esses empates já não são permitidos.

O método da categorização de Thurstone é de uso aconselhável, por exemplo, quando o número de estímulos é grande. Com efeito, a comparação por pares no modelo dos julgamentos comparativos exige de cada juiz a manifestação de $\frac{m(m-1)}{2}$ julgamentos, enquanto no modelo de categorização os julgamentos correspondentes são em menor número.

Pelas razões já consideradas anteriormente, tanto as diferenciações nos julgamentos dos juizes quanto as próprias variações temporais dos seus comportamentos levam agora à substituição dos valores de escala μ_i e C_j por variáveis aleatórias ξ_i, e η_j, nessa ordem, além, é claro, da mudança da regra decisòria para a sua correspondente expressão probabilistica na categorização. Nessas circunstâncias, tendo em vista o caráter classificatòrio do modelo, designar-se-á por $\Pi_{ij}^{(A)}$ a probabilidade de que o estímulo O_i seja classificado em uma das j primeiras categorias representadas pelos intervalos I_1,, I_2, ..., I_3, ou seja:

$$(5.10) \quad \Pi_{ij}^{(A)} = Prob[O_i \in I_1 \cup I_2 \cup \dots \cup I_j],$$

onde o superindice (A) designa o fato de que se trata de uma probabilidade acumulada. Do exposto, então, pode-se afirmar que tal probabilidade expressa-se em termos dos valores da escala aleatorizada, através da equação $\Pi_{ij}^{(A)} = Prob[\xi_i \leq \eta_j]$ (i = 1, 2,..., m; j = 1, 2,..., k)

Thurstone impõe, também às variáveis aleatórias ξ_i e η_j, além da normalidade de suas distribuições, as seguintes estatísticas:

i)($\xi_i = N(\mu_i, \sigma_i^2)$

ii) $\eta_j = N(c_j, \psi_j^2)$ (j=1,2,..., k)

iii) são nulas todas as correlações entre as variáveis aleatórias.

Tratando a última expressão da probabilidade acumulada do mesmo modo que no modelo dos julgamentos comparativos, pode-se concluir facilmente que

(5.11) $$\Pi_{ij}^{(A)} = Prob\left[\frac{(\xi_i-\eta_j)-(\mu_i-c_j)}{\sqrt{Var(\xi_i-\eta_j)}} \leq -\frac{(\mu_i-c_j)}{\sqrt{Var(\xi_i-\eta_j)}}\right]$$

Do mesmo modo, considerando que as variáveis aleatórias de escalagem têm distribuições normais, além de incorrelacionadas, pode-se escrever que

$$-\frac{(\mu_i-c_j)}{\sqrt{Var(\xi_i-\eta_j)}} = z_{ij},$$

ou seja: $c_j-\mu_i = z_{ij}\sqrt{Var(\xi_i-\eta_j)}$, ou ainda

$$c_j-\mu_i = z_{ij}\sqrt{\sigma_i^2+\psi_j^2}$$

Supondo que sejam constantes as variâncias $Var(\xi_i-\eta_j)$ e, portanto, sem perda de generalidade, fazendo-se $Var(\xi_i-\eta_j)=1$ (i = 1, 2, ..., m; j = 1, 2, ..., K), chega-se ao seguinte sistema de equações $c_j-\mu_i=z_{ij}$ (i = 1, 2, ..., m; j = 1,2,..., K), onde z_{ij} representa a abcissa da curva normal tal que a área à sua esquerda corresponde à probabilidade do estímulo O_i estar nos j últimos lugares de preferências dos juizes.
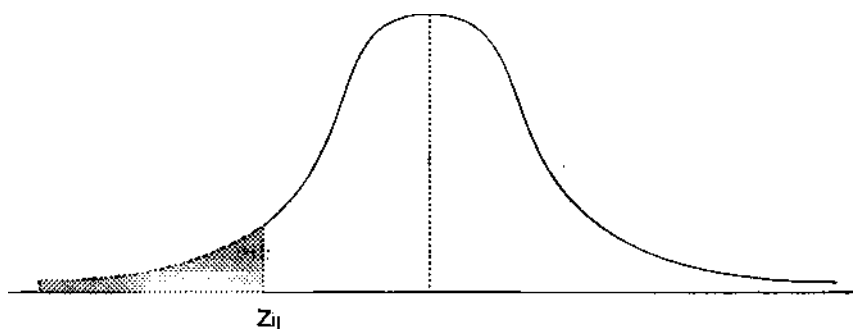
Fig. 5-2

A estimação mínimo-quadrática dos m + k parâmetros é objeto de minimização da função

Derivando-a relativamente a μ_i e c_j e escrevendo as equações de 1ª ordem de otimização, obtêm-se as equações:

$$\begin{cases} \sum_{j=1}^{k} -2[(\hat{c}_j - \hat{\mu}_i)] - z_{ij}] = 0 \\ \sum_{i=1}^{m} 2[(\hat{c}_j - \hat{\mu}_i) - z_{ij}] = 0 \end{cases}$$ (i = 1, 2,..., m; j=l,2, ...,k)

donde se conclui que

$$\begin{cases} \sum_{j=1}^{k} \hat{c}_j - k\hat{\mu}_i = z_{i.} \\ m\hat{c}_j - \sum_{i=1}^{m} \hat{\mu}_i = z_{.j} \end{cases}$$ (i- 1,2, ...,m; j= 1,2, ..., k)

ou ainda, equivocantemente,

$$\begin{cases} -k\hat{\mu}_i = z_{i.} - \sum_{j=1}^{k} \hat{c}_j \\ m\hat{c}_j = z_{.j} = \sum_{i=1}^{m} \hat{\mu}_i \end{cases}$$ (i- 1,2, ..., m; j= 1,2,..., k).

Convencionando também que o que não implica nenhu- $\sum_{i=1}^{m} \hat{\mu}_i = 0,$
ma perda de generalidade, obtém-se a expressão

(5.12) $\hat{c}_j = \bar{z}_{.j}$ (j = 1, 2..................... *k)*

Substituindo esse resultado nas primeiras equações tem-se que

$$-k\hat{\mu}_i = z_{i.} - \frac{1}{m}\sum z_{.j}$$, ou seja,

$$\hat{\mu}_i = \frac{1}{k} z_{i.} + \frac{1}{km} z_{..}$$ ou, ainda, que

$$\hat{\mu}_i = -\frac{1}{k} z_{i.} + \bar{z}_{..}$$ (*i* = 1.2........ *m).*

Tendo em vista a interpretação estritamente ordinal que se dá aos valores de escala dos estímulos, conclui-se que

$$\hat{\mu}_i = -\frac{1}{k} z_{i.}$$ (*i*= 1, 2.......... *m)* ou, finalmente,

(5.13)
$$\hat{\mu}_i = -\bar{z}_{i.} \quad (i = 1, 2, ..., m)$$

Ainda sob o mesmo enfoque metodológico dos modelos de Thurstone, pode-se desenvolver um outro modelo, devido a Luce (1963) e que possibilita aos juizes a natural opção por uma indiferença ou empate entre os estímulos, o que é muito freqüente no comportamento humano.

Considere-se, para esse fim, a correspondente extensão do modelo dos julgamentos comparativos de Thurstone ao caso em que os juizes podem optar por um empate entre as intensidades de dois estímulos. Desse modo, desde o ponto de vista decisional, aos valores de escala μ_i (1-1,2,..., m) acrescenta-se ao modelo de Luce um parâmetro "e" positivo (e > 0), de tal modo que os novos critérios de manifestação dos juizes são agora exprimidos do seguinte modo:

(5.14)
$$\begin{cases} O_i \rbrack O_j & \text{se e somen } \mu_i > \mu_j + e \\ O_i = O_j & \text{somente se } | \mu_i - \mu_j | \leq e \end{cases}$$

em que o símbolo « denota o empate entre os estímulos O_i e Oj e o parâmetro "e" positivo é uma espécie de limiar que, se não ultrapassado, impede a discriminação entre os estímulos O_i e *Oj* desde o ponto de vista da manifestação de uma preferência. Pela mesma ordem de idéias usadas anteriormente, transformam-se essas regras para a correspondente versão probabilistica, ou seja, impõem-se, agora, os seguintes critérios decisorios de caráter estatístico:

(5.15)
$$\begin{aligned} \Pi_{ij} &= Prob[O_i \rbrack O_j] = Prob[\xi_i > \xi_j + \Pi] = \\ &= \varepsilon_{ij} = Prob[O_i = O_j] = Prob[|\xi_i - \xi_j| \leq \Pi] \end{aligned},$$

onde n é uma variável aleatória normal com média igual a e e variância igual a ψ^2, e ε_{ij}, denota a probabilidade do empate ou indiferença O_i, = O_j.

Operando de modo análogo àquele do modelo de Thurstone, concebem-se as seguintes relações:

$$\begin{cases} \Pi_{ij} = Prob[\xi_i - \xi_j - \eta > 0] = \\ = Prob[(\xi_i - \xi_j - \eta) - (\mu_i - \mu_j - \theta) \geq (\mu_i - \mu_j - \theta)] \end{cases} \quad e$$

$$\begin{cases} \varepsilon_{ij} = Prob[-\eta \leq \xi_i - \xi_j \leq \eta] = Prob[r_i - \xi_j - \eta \leq \hat{0} r_i ; \xi_i + \eta \geq 0] \\ = Prob[(\xi_i - \xi_j - \eta) - (\mu_i - \mu_j - \theta) \leq -(\mu_i - \mu_j - \theta); \\ \qquad (\xi_i - \xi_j + \eta) - (\mu_i - \mu_j + \theta) \geq (\mu_i - \mu_j - \theta)] \end{cases}$$

(5.16)

(5.17)

Seja, agora, z_{ij} o valor da abcissa da curva normal (0,1) correspondente à preferência $O_i \rbrack O_j$, ou seja, a abcissa cuja área à sua direita, sob essa curva, é igual a Π_{ij}. Daí, vem que

(5.18)

$$z_{i|j} = \frac{(\mu_i - \mu_j - \theta)}{\sqrt{Var(\xi - \zeta - \eta)}}.$$

A preferência inversa O_i] O_j, por outro lado, fica caracterizada pela correspondente equação

Fazendo agora

$$= Pro\ldots \quad (\mu_i - \mu_j + \theta)]$$

(5.19)

$$z_{i[j} = \frac{-(\mu_i - \mu_j - \theta)}{\sqrt{Var(\xi_i \xi_j + \eta)}},$$

onde $z_{i[i}$ é a abcissa da curva normal (0,1) cuja área à esquerda desse valor é igual a $n_{ji;}$ deve-se notar que, desse modo, para o modelo Luce, encontram-se duas abcissas, de acordo com a figura 5-3.
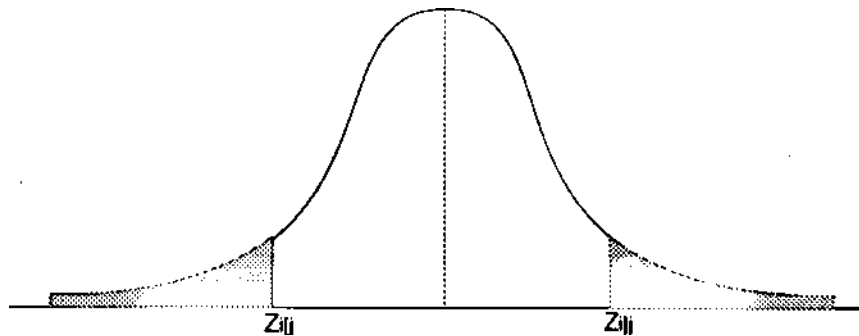
Fig. 5-3

Supondo agora, tal como no modelo de Thurstone, inter-relacionadas todas as variáveis aleatórias, torna-se claro que se tem

$$Var(\xi_i - \xi_j - \eta) = Var(\xi_i - \xi_j + \eta) = Var(\xi_i - \xi_j) + Var(\eta)$$

e, assim, pode-se escrever que

(5.20)

$$\frac{\mu_j - \mu_i + \theta}{\sqrt{Var(\xi_i - \zeta_j) + Var(\eta)}} = z_{i]j}$$

(5.21)

$$\frac{\mu_j - \mu_i - \theta}{\sqrt{Var(\xi_i - \zeta_j) + Var(\eta)}} = z_{i|j}$$

Supondo, também, simplificadamente, que as variâncias $Var(\xi_i - \xi_j)$ sejam independentes dos estímulos, concebe-se que $Var(\xi_i - \xi_j) + Var(\eta)$ será constante e, desse modo, pode-se fazê-la unitária porque esta hipótese corresponde apenas a uma mudança de escala nos valores de escala, sempre interpretados na escala ordinal. Desse modo, somando e subtraindo as duas equações anteriores, obtêm-se as expressões

$$\begin{cases} 2(\mu_j - \mu_i) = z_{i]j} + z_{i[j} \\ 2\theta = z_{i]j} - z_{i[j} \end{cases}$$

Diante das manifestações colhidas de n juizes, escrevem-se as seguintes expressões para os estimadores naturais ū, e ê (i = 1, 2,..., m), dos parâmetros do modelo:

(5.22)

$$\hat{\theta} = \frac{z_{i]j} - z_{i[j}}{2} \quad e$$

(5.23) $$\hat{\mu}_j - \hat{\mu}_i = \frac{z_{i]j} + z_{i[j}}{2}$$

Impondo, agora, do mesmo modo que antes aos estimadores ū, (i= 1,2,..., m) a condição Σū = 0, chega-se facilmente à expressão final

(5.24) $$\hat{\mu}_j = \frac{1}{2m}\sum_{i=1}^{m} z_{i]j} + z_{i[j}$$

O MODELO DE BRADLEY-TERRY

Bradley e Terry (1952) consideraram uma outra linha de argumentação para escalar um conjunto de m estímulos ou objetos psicológicos baseados, ainda, no mesmo método das pareações, ou seja, na comparação de todos os pares de estímulos efetuada por n juizes. Eles supuseram, para isso, que a cada estímulo O, está associado um valor de escala $\Pi_i > 0$ (i - 1, 2, ..., m) e que, para um juiz genérico, a probabilidade da preferência O,] O_j é definida pela equação

(5.25) $$Prob[O_i]O_j] = \frac{\Pi_i}{\Pi_i + \Pi_j} \quad (i \neq j)$$

Ressaltam, dessa hipótese, as duas seguintes conclusões:

I) a preferência porventura existente entre dois estímulos não é afetada pelos (m - 2) estímulos restantes;

II) a hipótese formulada assemelha-se a um modelo de urnas para a decisão sobre as preferências no *continuum* psicológico.

Ambas as restrições são motivos de fortes questionamentos que enfraquecem o embasamento teórico do modelo de Bradley-Terry. Entretanto, por sua simplicidade e, também, porque suas soluções ordenam os estímulos de modo muito assemelhado ao do modelo dos julgamentos comparativos de Thurstone, recomenda-se o seu uso.

No concernente à estimação de seus parâmetros, dois métodos podem ser chamados para esse fim. O primeiro, o método dos momentos, considera as freqüências absolutas n_{ij} observadas para as preferências $O_i \} O_j$. Por elas pode-se escrever, de modo natural, as equações que determinam os estimadores resultantes do método dos momentos

$$(5.26) \quad \frac{\hat{\Pi}_i}{\hat{\Pi}_i - \hat{\Pi}_j} = \frac{n_{ij}}{n_{ij} + n_{ji}} \quad (i \neq j),$$

sem que, necessariamente, se tenha $n_{ij} + n_{ji} = n$, ou seja, é possível não opinar quanto à preferência no modelo proposto.

Após manipulações algébricas elementares, esse sistema se transforma no sistema linear e homogêneo escrito matricialmente como

$$(5.27) \quad \begin{bmatrix} -n_{.1} & n_{12} & n_{13} & \cdots & n_{1m} \\ n_{22} & -n_{.2} & n_{23} & \cdots & n_{2m} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ n_{m1} & n_{m2} & n_{m3} & \cdots & -n_{.m} \end{bmatrix} \begin{bmatrix} \hat{\Pi}_1 \\ \hat{\Pi}_2 \\ \vdots \\ \hat{\Pi}_m \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

ou, ainda, equivalentemente,

$$(5.28) \quad N\hat{\Pi} = 0$$

É fácil ver que a matriz N é singular porque suas linhas são linearmente dependentes (são nulas as somas dos elementos de cada coluna), e que, por outro lado, os valores de escala devem satisfazer à condição natural

$$\hat{\Pi}_1 + \hat{\Pi}_2 + \dots + \hat{\Pi}_m = 1$$

Desse modo, substituindo uma equação genérica do sistema (5.28) por essa última equação, pode-se encontrar uma solução do problema. Assim, sendo substituída a linha j da matriz N pelo vetor 1 e sendo e_j o j-ésimo vetor unitário, ou seja, o vetor cuja j-ésima componente é igual à unidade sendo nulas as demais, o sistema de equações anterior é expressado matricialmente por

$$(5.29) \quad N_j\hat{\Pi} = e_j,$$

onde N_i é nova matriz que resulta de N pela substituição antes enunciada. Nesse caso, a solução do problema de escalagem é igual a

$$\hat{\Pi} = N_j^{-1} e_j$$

(5.30)

ou seja, fi é a j-ésima coluna da matriz N_i^{-1}, inversa de N_i, Pode-se mostrar, facilmente, que essa solução independe de j, ou seja, que ela é independente da equação substituída.

O outro método de estimação para o modelo de Bradley-Terry segue um algoritmo geral de Ford (1957), desenvolvido para resolver um sistema de equações resultante da aplicação do método da máxima-verossimilhança. Assim, para as observações do modelo de Bradley-Terry é fácil concluir que sua função de verossimilhança é dada pela expressão

(5.31) $$L = \prod_{i<j} \left[\frac{\Pi_i}{\Pi_i + \Pi_j}\right]^{n_{ij}} \left[\frac{\Pi_j}{\Pi_i + \Pi_j}\right]^{n_{ji}}$$

Convencionando-se que $n_{ii} = 0$ (i = 1,2, ..., m) e maximizando o logaritmo natural Lg L da função de verossimilhança, obtém-se facilmente o seguinte sistema de equações máximo-verossímeis, cuja solução fornece os valores de escala dos estímulos

(5.32) $$\begin{cases} \dfrac{n_{i.}}{\hat{\Pi}_i} - \displaystyle\sum_{j, j \neq i} \dfrac{n_{ij} + n_{ji}}{\hat{\Pi}_i + \hat{\Pi}_j} = 0 \quad (i = 1, 2, ..., m) \\ \hat{\Pi}_1 + \hat{\Pi}_2 + ... + \hat{\Pi}_m = 1 \end{cases}$$

Dessas equações resulta que

(5.33) $$\hat{\Pi}_i = \frac{n_{i.}}{\displaystyle\sum_{j, j \neq i} \frac{n_{ij} + n_{ji}}{\hat{\Pi}_i + \hat{\Pi}_j}}$$

O método de solução de (5.33) consiste em usar um algoritmo iterativo convergente usando a seguinte fórmula de recorrência, em que

$$\hat{\Pi}_i^{(k+1)} = \frac{n_{i.}}{\displaystyle\sum_{j, j \neq i} \frac{n_{ij} + n_{ji}}{\hat{\Pi}_i^{(k)} + \hat{\Pi}_j^{(k)}}} \quad (k = 1, 2, ...)$$

$\Pi^{(k)}$ denota a aproximação de Π_i, na k-ésima iteração do algoritmo

(5.34)

A implantação do algoritmo anterior requer, para sua implementação mais eficaz, as seguintes providências:

I) uma boa aproximacão inicial para a sua rápida convergência;

II) uma norm $\hat{\Pi}_i^{(k+1)}$ (i = 1, 2, ..., m) ações obtidas em cada etapa.

A aproximação inicial, aqui sugerida, consiste simplesmente em fazer $\hat{\Pi}_i^{(1)} = \frac{1}{-}$ (i = 1,2,..., m), isto é, supõe-se inicialmente uma preferência manifestada imparcialmente sob a hipótese de absoluta ignorância no concernente à natureza dos estímulos.

A condição de normalização é naturalmente imposta em cada etapa e consiste em substituir cada $\hat{\Pi}_i^{(k)}$ por

$$\frac{\hat{\Pi}_i^{(k)}}{\sum_{j=1}^{m} \hat{\Pi}_j^{(k)}}$$

para satisfazer à condição $\sum_{i=1}^{m} \Pi_i^{(k)} = 1$.

É possível considerar, também, como Rao and Kuper (Sijberg, 1967), uma generalização do modelo de Bradley-Terry para escalagem de estímulos de modo a contemplar os julgamentos de preferências que passem pela indiferença entre dois estímulos. O m...óprio consiste, obviamente, em reduzirem-se as probabilidades $\frac{\Pi_i}{\cdots}$ das preferências O_i] O_p de modo a poder debitar seus saldos nas indiferenças $O_i = O_j$.

Segundo esse enfoque, agrega-se ao modelo tradicional de Bradley-Terry um parâmetro 0 (O<0<1), de tal sorte que se tenham as seguintes probabilidades:

$$Prob[O_i] O_j] = \frac{\Pi_i}{\Pi_i + \Theta\Pi_j} \tag{5.34}$$

$$Prob[O_i] O_j] = \frac{\Pi_j}{\Pi_j + \Theta\Pi_i} \tag{5.35}$$

$$Prob[O_i \approx O_j] = 1 - \left[\frac{\Pi_i}{\Pi_i + \Theta\Pi_j} + \frac{\Pi_j}{\Pi_j \Theta\Pi_i}\right] = \frac{(\Theta^2 - 1)\Pi_i\Pi_j}{(\Pi_i + \Theta\Pi_j)(\Pi_j + \Theta\Pi_i)} \quad (i \neq j) \tag{5.36}$$

A estimação, nesse caso, também pode ser submetida ao mesmo algoritmo descrito anteriormente.

O leitor deve estar atento a algumas diferenciações entre as duas grandes linhas clássicas de desenvolvimento de modelos de escalagem psicológica unidimensional. Cumpre, assim, em primeiro lugar, dar destaque ao fato de que os modelos dos julgamentos comparativos e dos julgamentos categóricos de Thurstone são dotados de uma racionalidade aceitável desde o ponto de vista da teoria psicológica. Esta racionalidade, entretanto, já não é aceitável pelo modelo de Bradley-Terry. Este último assemelha-se a certas formulações da teoria matemática da aprendizagem, desde que se concebam a irracionalidade do processo decisòrio e a contrapartida de penalidades ou sanções para uma escolha equivocada. Nessa situação, é óbvio, as opções de preferências entre os estímulos Θ_i e Θ_j, deverão repartirem-se proporcionalmente às probabilidades $\frac{\Pi_i}{\Pi_i + \Pi_j}$ e $\frac{\Pi_j}{\Pi_i + \Pi_j}$.

Além disso, no que se refere, ainda, às concepções psicofísicas de Thurstone, cujas origens remontam aos trabalhos de Fechner na escola alemã de psicofisica, já foi feita referência à influência que sobre suas idéias exerceu a teoria dos erros, que durante muito tempo dominou as formulações estatísticas de Pearson, de Fisher e de Neyman da teoria estatística clássica. Nada impede, entretanto, que as variáveis aleatórias de escalagem tenham distribuições assimétricas representando as disposições psicológicas diferentes dos juizes que, no caso de assimetrias, acolhem as naturais tendências de se fazer avaliações severas ou magnânimas.

Com efeito, um juiz severo tenderá a fazer avaliações baixas em detrimento das avaliações mais elevadas e, contrariamente, um juiz magnânimo se comportará seguindo uma postura inversa. Nessas circunstâncias, torna-se perfeitamente admissível supor que as variáveis aleatórias de escalagem ξ, dos estímulos O_i, tenham distribuições de probabilidades F(x) assimétricas ou concentradas lateralmente.

Para obter a generalização intuída a partir dos comentários anteriores, considere as equações de escalagem do modelo dos julgamentos comparativos de Thurstone, $-(\mu_i - \mu_j) = z_{ij}$ $(i \neq j)$.

Neste caso, uma outra forma de expressá-las consiste em escrever que

$$\Pi_{ij} = Prob[O_i]O_j] = \int_{-(\mu_i-\mu_j)}^{+\infty} \frac{1}{\sqrt{2\Pi}} e^{-\frac{x^2}{2}} dx \quad (i \neq j) \tag{5.37}$$

ou, ainda, que

$$
\begin{aligned}
\Pi_{ij} &= 1 - F[-(\mu_i - \mu_j)] \quad (i \neq j) \\
\Pi_{ij} &= 1 - F(\mu_j - \mu_i) \quad (i \neq j)
\end{aligned}
\tag{5.38}
$$

Segundo esse resultado, ao utilizar-se, por exemplo, uma distribuição do tipo gama, apresentam-se ao modelo os seus dois parâmetros caracterizadores, além, é claro, dos valores de escala dos estímulos. Esses parâmetros, quando estimados, permitem fazer apreciações sobre a severidade ou a magnanimidade dos juizes.

Desde um outro ponto de vista, entretanto, a dicotomia sobre a justeza dos julgamentos efetuados pode ser o fruto exclusivo das disparidades nos conteúdos latentes dos estímulos examinados, ou seja, se a maioria deles contém baixos teores do conteúdo latente e os juizes são equilibrados, a tendência destes é atribuir-lhes valores de escala concentrados à esquerda, enquanto, no caso inverso, os valores de escala concentrar-se-ão à direita. Estas observações apontam para novas direções no exame do tema.

REFERENCIAS BIBLIOGRÁFICAS

BRADLEY, R.A.. TERRY. M.E. The rank analysis of incomplete block designs I: the method of paired comparisons. *Biometria,* v.39, p.324-345, 1952.

FORD JR., L.R. Solution of a ranking problem from binary comparisons. *American Mathematical Monthly.* Washington. D.C.. v.64. p.28-33, 1957.

GREENBERG, MG. A modification of Thurstone's law of comparative judgment category of equal or no difference. *Psychological Bulletin.* Washington. D.C.. v.64. 108-112, 1965.

GUTTMAN, L. An approach for quantiting paired comparisons and rank order. *Annals of Mathematical Statistics.* Hayward (CA), v.17, p.144-163. 1946.

HOTELLING, N. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology.* Washington. D.C.. v.24, p.417-441; 498-520,1933.

LUCE, R.D. Discrimination. In: LUCE, R.D., BUSH, R.R., GALANTER, E. *Psychology.* New York: Wley. 1963. v.1.

MICHAUD. P. Opinions aggregations. In: JANSE, J.. MARCOTORCHINO, J.F., PROTH, J. M. (Eds.). *New trends in data analysis and applications.* Amsterdam: North Holland, 1983.

NISHISATO, S. *Analysis of categorical data:* dual scaling and its applications. Toronto: University of Toronto Press, 1980.

---------------. Optimal scaling of paired comparison and rank order data: an alternative to Guttman's formulation. *Psychometrika,* v.43, p.263-271, 1978.

SIJBERG, L. Successive intervals scaling of paired comparisons. *Psychometrika,* v.32, p.297-308,1967.

SOUZA, J. de. *Métodos de escalagem psicossocial.* Brasília: Thesaurus, 1987.

THURSTONE, L.L. A law of comparative judgment *Psychological Review.* Washington, D.C., v.34, p.273-286, 1927.

CAPITULO 6

EM TORNO DO ENSINO DE ESTATÍSTICA NOS CURSOS DE BASE PSICOSSOCIAL

Jorge de Souza
Departamento de Economia
Universidade Federal de Pernambuco

É mais do que freqüente — e eu diria mesmo justificável — a rejeição por psicólogos, sociólogos e outros profissionais aos métodos psicossociométricos e às muitas técnicas estatísticas abusiva e equivocadamente aplicadas no trabalho investigativo nos domínios de suas ciências. Causas diversas contribuem para essa verdadeira aversão aos conhecimentos psicométricos e estatísticos, que se manifesta, inicialmente, por um difuso sentimento de inadequação de suas formulações, sobretudo no que concerne à ciência psicométrica fundada pelos notáveis estatísticos ingleses Charles Spearman e Karl Pearson, no que diz respeito aos seus fundamentos estatísticos, e pelo médico e psicólogo francês Alfred Binet, no que se refere à formulação adequada dos instrumentos conceituais de mensuração psicológica."

De fato, essa inadequação generalizadamente sentida, mas insatisfatoriamente justificada, é, em muitos casos, o resultado de concepções teóricas que não encontram respaldo nem nas teorias estabelecidas e muito menos na experiência com as fenomenologías psicológica, sociológica ou psicossocial, de um modo mais amplo. É isso que se dá, por exemplo, no uso abusivo que se faz da clássica Análise Fatorial com fatores ortogonais como instrumento de pesquisa no domínio do psicossocial. Atropela-se com ela qualquer expressão de bom-senso, pela impertinencia do uso desrespeitoso dos fundamentos dessa técnica tão atraente quanto perigosa.

Esse exemplo serve ao propósito de considerar a questão objeto deste capítulo por suas verdadeiras dimensões epistemológicas. Com

* Note-se que. neste trabalho, concebem-se de modo inovador as bases históricas da Psicometria como o resultado da confluência entre os esforços de psicólogos e estatísticos. Para um mais amplo entendimento de uma formulação da história da Psicometria, pode-se consultar, deste mesmo autor. *História das Idéias Estatísticas.* Recife: UFPE. 1995. Notas internas.

efeito, no caso em tela, a inadaptabilidade do método fatorial manifesta-se, em primeiro lugar, pela justificável rejeição que se deve esperar relativamente a uma das hipóteses nele implícitas, configurada como a incorrelação dos fatores ou variáveis latentes. Nesse sentido, o leitor já deve beneficiar-se da dúvida natural que se lhe acomete — sub-repticiamente, é bem verdade — quanto à adequação da Análise Fatorial. Mas essa mesma rejeição aos fatores ortogonais pode ser também justificada pela ignorância de quem os manuseia no referente à hipótese de ortogonalidade ou incorrelação entre os fatores.

Infere-se daqui, na realidade, uma dupla manifestação da ignorância científica. Em primeiro lugar, no concernente às hipóteses fundamentais dessa fértil técnica psicométrica e, em segundo lugar, pelo desconhecimento de alternativas metodológicas que superam satisfatoriamente a restrição interposta. Assim, no caso, poder-se-ia optar pelo método sucedâneo da Análise Fatorial Oblíqua como um meio propício à superação da séria restrição de ortogonalidade. Esta, sabe-se bem, não encontra respaldo nem nos dados observacionais e nem, tampouco, nas formulações teóricas subjacentes de todas as ciências psicossociais — pode-se dizer sem medo de errar.

Não se há de inferir, desse exemplo — e essa formulação consubstanciaria, também, um erro inaceitável —, que a Análise Fatorial clássica com fatores ortogonais deva ser abandonada enquanto conhecimento metodológico — não. O leitor mais versado nas literaturas estatística e psicométrica iria logo nelas reconhecer a sua imprescindibilidade como base ou etapa intermediária para alcançar a formulação mais pertinente ao embasamento teórico da sucedânea — e mais pertinente — Análise Fatorial Oblíqua.

Toda essa teia de considerações apoiada no paradigma da Análise Fatorial auxilia-nos a concluir que, no geral, é ignorância quanto às bases ou hipóteses que sustentam os métodos ou as técnicas psicométricas e estatísticas o fator primordial de sua inadequação na abordagem dos dados observacionais. Na realidade, essa última afirmação é por demais abusiva, porque a inadaptabilidade constatada não reside nos métodos em si mesmos, mas nas suas condições de aplicação. É que eles se comportam com neutralidade, esperando, ingenuamente, que seus usuários saibam manejá-los com respeito às suas idiossincrasias. Seria fácil e cientificamente desonesto, convenhamos, assacar contra a Psicometria ou a Estatística, rés inocentes que padecem da mesma injustiça com que as uvas maduras, na fábula de La Fontaine, passaram por verdes pela frustração da raposa incapacitada para alcançá-las.

Uma outra ordem de idéias sobre a insuficiência dos conhecimentos psicométrico e estatístico resulta da consideração dos fins a que eles se destinam. Para ilustração, ponha-se, ainda, o exemplo com que se vem

tratando o tema. E muito comum encontrarem-se dezenas de complexos artigos que versam sobre a inferência estatística em Análise Fatorial. A tão conhecida e importante revista pioneira *Psychometrika* é detentora de recordes nesse sentido. Muito bem — digo com ênfase —, esses artigos laureiam seus autores pela originalidade de suas concepções, pela sofisticação estatística e matemática que nos deixam complexados e com sentimentos mesmo que ora são os da inferioridade, vezes outras os da frustração e — esse é o mais justo, reconheçamos — o do temor! Todas essas dolorosas sensações, infelizmente, não são manifestadas de modo explícito; não as revelamos para não nos desnudarmos diante dos colegas que, eles mesmos, ironicamente, também as padecem no altar sofrido de limitações iguais às que nos acicatam. E, se algo revelamos sobre isso em algum momento, é uma espécie de rancor que gera, a meu juízo, esse sentimento difuso de aversão à Psicometria e, também, à Estatística de um modo geral. Como bom psicanalista inquieto, eu veria como normais esses sentimentos e, até mesmo, considerá-los-ia mais que pertinentes ou saudáveis. É que há algo de errado ou talvez inadequado por trás dessas formulações estéreis, e é exatamente essa sensação que quero analisar friamente. Onde sustentar-se, então, para considerá-la em suas dimensões corretas? Em que se podem apoiar os mortais psicólogo e sociólogo, condenados ambos, juntamente com politicólogos e pedagogos, ao sacrifício irremediável nesse altar de formalismos lógico-matemáticos? Há um caminho fácil, leitor, quero dizer, perfeitamente trilhável, porque possui para o intento apenas a difícil e delicada capacidade de observar; mas deve-se fazê-lo, é claro, judiciosamente e sem idéias preconcebidas — isso é essencial. Com efeito, todos esses conhecimentos — que se diriam de cunho estatístico-inferenciais relativamente à Análise Fatorial — sustentam-se, entre várias outras, na hipótese simples de que os erros dos testes ou instrumentos de medição que geram os escores ou notas têm distribuição normal de Gauss-Laplace. Essa é a questão que deve ser considerada quando nos propomos a dissecar a plausibilidade de Psicometria no contexto psicossocial a que ela se destina. Interroguemonos, assim, se a destacada hipótese estatística sobre a normalidade dos erros decorre da observação, do exame dos fatos, ou se, contrariamente, ela é fruto da gratuidade ou da conveniência com que, Procustos neuróticos da novidade científica, limitamos a realidade ao leito de ferro de nossas conveniências acadêmicas, para brilharmos nesse mundo falso de publicações e bibliografias!

Não há aqui como fugir-se, também, à sempre presente historicidade da ciência, à sua realidade de época, às marcantes influências importadas e, sobretudo, engolidas sem a imprescindível digestão presidida pelo amparo da crítica epistemológica! Com efeito, a Análise Fatorial, desde os seus primórdios londrinos de 1904, sob a inteligência de Spearman e, sobremodo, sob a tutela posterior de estatísticos como Hotelling, Thurstone, Wishart, Roy, Fisher e outros que, nos anos 30 e 40

deste século, consolidaram-na como uma teoria normal da Analise Estatística Multivariada, estava marcadamente influenciada pela mais que centenária teoria dos erros normais, essa fecunda fonte de toda ciência observacional que sustentou, também, a Estatística Inferencial da linha desenvolvida de Ronald Fisher, Egon Pearson e Jerzy Neyman. No entanto, é preciso que se ressalte, essa hipótese de normalidade, descoberta dos astrônomos e físicos, era adequada aos dados das ciências experimentais, era ditada pelas observações das ciências cujos experimentos eram controlados pelo pesquisador. Diante disso, a conclusão que nos compete enunciar é a de que, infelizmente, o conceito de erro distribuído normalmente não é respaldado pela natureza das observações das variáveis psicossociais obtidas, desde logo, mediante instrumentos de mensuração mais precários de que os das ciências experimentais. Adotá-la, assim, seria amarrar a realidade ao mesmo leito de ferro das conveniências que nos servem ao papel de Procustos da ciência e não a de verdadeiros cientistas ou, pelo menos, a de profissionais honestos e capacitados. Que fazer? Essa é a interrogação que, estarrecidos justificadamente, nos fazemos agora. Não há como precipitar-se diante disso. Longe, muito longe mesmo, a idéia de atirar tudo a esse lixo da ciência, tão impressionantemente rico de teorias, métodos e técnicas exibidos pretensiosamente como os mais puros produtos da inteligência e da cultura humanas. Contenhamos a indignação natural diante da descoberta de que tínhamos razão, de que não éramos tão tolos quanto as complicações estatísticas nos faziam crer! Já não nos devemos vexar, mas ver a tudo com reservas deve ser a primeira atitude, o que sugere a prudência, boa guia dos nossos propósitos. E mais: antes de quaisquer outras considerações, havemos de concluir que a posição mais correta é a de que, no estágio em que se situa a inferência da Análise Fatorial — e é sobre ela que se sustenta o nosso paradigma —, é que não convém adotá-la a não ser em situações muito específicas. Isto porque, antes de mais nada, é imprescindível ver a Análise Fatorial Clássica como um método eminentemente exploratório, sem nenhum cunho inferencial ou de confirmação de hipóteses, que, este sim, é o cerne da Inferência Estatística. Aliás, quase todo o aparato psicométrico até hoje conhecido convém que seja visto muito mais pela ótica exploratória ou descritiva e, menos do que desejam os seus teóricos desavisados e usuários incautos, pela visão ideal da indução estatística.

A conclusão a que chegamos sobre tudo o que se disse é a de que, no geral, o insucesso e a suspeição com que são vistos os métodos psicométricos e estatísticos são fruto de várias causas que se somam nessa empreitada de desprezo pelo bom-senso: a falta de conhecimento das hipóteses estatísticas que lhes dão respaldo, o seu uso abusivo, extrapolando-se os limites sem que para isso haja autorização científica e técnica e, *last but not least,* o desconhecimento de alternativas metodológicas mais consentâneas.

A Psicologia se insere no frondoso ramo das ciencias psicossociais, ou seja, no elenco de tôdas as ciências que lidam com o subjetivismo psicológico do homem e das relações humanas. Esse fato tem, para o estatístico, uma conotação muito especial porque, contrariamente às ciências experimentais, as ciências psicossociais, de que a Psicologia e a Sociologia são membros proeminentes, são preponderantemente dominadas por algumas características que, tornando-as especiais, implicam uma fundamental diferenciação metodológica do tratamento estatístico de seus problemas. Com efeito, apesar de que, em alguns contextos, possam a Psicologia e a Sociologia em particular submeterem-se ao método experimental, na quase absoluta maioria das vezes elas são objeto do chamado método observacional. Dessa importante distinção de cunho metodológico o que mais ressalta é a grande impossibilidade de exercício do controle direto de fontes de variação, ou seja, a sustentação experimental da homogeneidade ou invariância de certos fatores causais, tão típicos das ciências experimentais. É esse mesmo controle junto com o princípio de aleatorização, sabe-se bem, que vão propiciar a fundação, pelo notável estatístico inglês, Sir Ronald Aylmer Fisher, do Delineamento Experimental e da Análise Estatística da Variância. Esses métodos estatísticos propiciaram um grande avanço à experimentação, não só por dotá-la de instrumentos práticos de análise como por possibilitar-lhe a própria cidadania científica.

A distinção mais marcante entre as ciências observacionais e as experimentais, todavia, dá-se a partir de outras características que separam integralmente as abordagens estatísticas de uma e de outra. É desse modo que, por exemplo, tem-se prevalência nas ciências psicossociais das variáveis definidas por construtos ou traços. Um construto, sabe-se bem, é um modo abstrato de conceber uma variável correspondente a uma grandeza que não pode ser medida diretamente. São exemplos de construtos a inteligência, a neurose e a ideologia. Um construto ou traço também recebe outras denominações, conforme o domínio de conhecimentos onde ele se insere. Dessa forma, são sinônimas de construto as expressões variável latente, variável abstrata e fator, tomadas da própria Psicologia ou da Sociologia. Pode-se dizer, além disso, que um construto também se representa pela definição operacional, mas o fato que se deseja pôr em relevo nesta abordagem é a sua marcante presença nas ciências psicossociais em contraposição à sua quase total ausência nas ciências experimentais. Enquanto nessas últimas a mensuração é o resultado da aplicação de um instrumento físico confiável, adredemente preparado para esse clássico fim, nas primeiras a mensuração resulta do uso de um instrumento conceptual, de confiabilidade duvidosa e fundamentalmente de validez discutível. Com efeito, se, por exemplo, quer-se usar um teste psicológico para medir a inteligência ou um questionário

para avaliar o aprendizado, não há grande precisão ou estabilidade estatística nas correspondentes avaliações e, com muito menos razão ainda, pode-se garantir que eles medem, de modo incorruptível, os construtos para os quais foram destinados. Tem-se o direito de perguntar, em um questionamento natural a respeito e abrindo-se alternativas a muitas dúvidas, se efetivamente o teste psicológico mede a inteligência ou a cultura e se o questionário mede o aprendizado ou a memória... Os problemas da confiabilidade e da validez assumem, assim, um dramático papel nas ciências psicossociais, papel esse praticamente inexistente nas ciências experimentais, pelo menos no que tange à validez dos seus instrumentos de mensuração. Ninguém duvida, por exemplo, que um termômetro sirva a outros propósitos que não o de medir temperaturas. Entretanto, essa mesma certeza não preside o uso de um teste de inteligência, para ficar apenas nesse caso mais facilmente perceptível.

Emergem dessas considerações, também e com destaque, a dramática importância dos erros de mensuração, além da multiplicidade de formas de medir a mesma variável latente. Esta última questão não ocorre na Física e nem tampouco na Química, por exemplo, mas, nas ciências psicossociais é comum lidar-se com ela sob a denominação de Problema dos Indicadores Múltiplos. Só agora, mais recentemente, através dos chamados Modelos Estruturais Lineares (LISREL) ou Modelos de Estruturas de Covariancia, é que psicometristas, sociometristas, politicólogos e econometristas têm atentado para essa importante questão. É assim, em decorrência dessas abordagens mais recentes, que já se pode falar, para o espanto dos economistas, em uma Econometria velha e em uma Econometria nova!

Há, no entanto, uma outra característica que também distingue profundamente, do ponto de vista da análise estatística, as duas grandes ramificações das ciências objeto destas digressões. Queremos referir-nos, nesse contexto, às escalas de mensuração ou de expressão das variáveis. Com efeito, enquanto nas ciências experimentais é pequena a presença e o uso das escalas nominal e ordinal, nas suas pobres coirmãs, as ciências psicossociais, elas são preponderantes e imprescindíveis. Qual o significado estatístico dessas constatações que eu enfatizo como o nó górdio do ensino da Estatística e da Psicossociometria nos cursos de Psicologia, Sociologia, Ciências Políticas e Pedagogia? A resposta inicial e que vai surpreender, creio eu, a maioria dos leitores é a quase absoluta ausência de conteúdos estatísticos correspondentes a essas necessidades, típicas das ciências psicossociais, nos respectivos cursos universitários. Pode-se propor essa mesma questão de uma outra forma ainda mais contundente. Com efeito, pergunta-se: no estudo da Psicologia, por exemplo, onde há uma maciça presença de variáveis nominais e ordinais, por que o usual tratamento estatístico dessas variáveis não vai além de uma descobrida e pouco reveladora representação gráfica ou do cálculo

de um sumário modal, até o atrevimento de uma mais ousada medida de associação estatística? E esta, convenhamos, não vai mesmo além do coeficiente de associação de Goodman-Kruskal e jamais prescinde do sempre mal utilizado e pouco compreendido qui-quadrado de Pearson? E as inferências estatísticas, ensinadas temerariamente já num primeiro curso de graduação e a alunos absolutamente jejunos de qualquer compreensão da problemática científica, são pertinentes às necessidades profissionais da Psicologia? E os sociólogos, o que andam estudando? E mais ainda, pode-se perguntar: os problemas científicos das ciências de teor psicossocial, que nos servem ao exercício crítico neste trabalho, são compatíveis com as inexoráveis hipóteses estatísticas subentendidas pelos métodos tão irresponsavelmente ensinados? A resposta a essas questões não pode prescindir do registro do espanto com que eu vislumbro as faces dos meus leitores. Há algo de podre nesse reino do ensino da Estatística e da Psicometria nas ciências psicossociais, e nós todos, estatísticos ou psicossociometristas, não nos fartamos do neologismo, temos alimentado irresponsavelmente esse estado de coisas, seja por omissão, por falta de conhecimento ou, ainda, pela ausência de um exercício crítico que a pressa na neurótica busca de resultados novidadosos nos tem eximido de fazer.

Há uma outra característica que distancia as ciências psicossociais das ciências experimentais. Quero referir-me, aqui, à multidimensiona-lidade que marca as primeiras e que se acha bem menos representada nessas últimas. Com efeito, os fenômenos comportamentais de base psicológica requerem, para seu exame, a consideração de muitas variáveis que se lhe associam numa interação ou sinergia que tornam não só difícil o isolamento de uma ou de poucas entre elas, mas que fazem prevalecer o conceito de associação ou de correlação estatística diante da noção clássica de causalidade. Emergem dessas considerações o caráter mul-tivariado e acausai das variáveis e fenômenos psicossociais que, por sua vez, vão fazer prevalecer sobre a sua abordagem estatística a chamada Análise Multivariada. Entretanto, o que se ensina aos profissionais desses domínios é, preponderantemente, Estatística Univariada ou Unidimensional que, é óbvio, não atende às suas necessidades.

Eu desejo enfatizar, também, para que não vistamos sozinhos a carapuça de desavisados, que os problemas até aqui exibidos são universais. Não se trata, dessa maneira, de circunscrevê-los a uma esfera geográfica terceiro ou primeiro-mundista e nem tampouco ocorre unicamente na formação básica dos cientistas e profissionais da Psicossociologia em detrimento da correspondente educação pós-graduada. Todos são padecentes desse grande engodo que tem sido o ensino da Estatística.

Nesta altura, entretanto, já não há como prescindir do enunciado de algumas causas, sob pena de estar aqui, este autor, a despejar nos leitores umas tantas gratuidades, porque, reconheço também, de equívocos

e certas exagerações que fortificam do gosto literário, mas auxiliam-se na pedagogia, não me eximo. Haverei de cometê-los sim, mas no grau menor do passável, do não-essencial, e isso é o que importa. Há algumas respostas — volto às indagações para que possamos entender esse estado de coisas — e cito três delas, todas convergentes e as mais importantes a meu juízo. A primeira é a da tradição do ensino não-crítico dos métodos estatísticos, que foram desenvolvidos e sistematizados, com sucesso, como já se disse, sob a esfera das demandas exigentes das ciências experimentais; a segunda é a da falta de sistematização que tem marcado o desenvolvimento dos métodos estatísticos e psicométricos apropriados ao tratamento dos problemas da alçada psicossocial; e a terceira, finalmente, é a ausência de livros-textos críticos e atualizados, o que tem possibilitado esse verdadeiro estelionato estatístico com que a boa-fé dos nossos educandos tem sido desgraçadamente ludibriada.

Há, também, unindo-se a essas causas, uma força de inércia que age não somente sob o domínio exclusivo das ciências psicossociais, mas é geral no concernente ao ensino universitário da Estatística. Com efeito, como responder a essa verdadeira obsessão de dotar os profissionais de qualquer ramo da ciência de conhecimentos dos dificílimos métodos estatísticos? Por que o sociólogo deve saber medir e por quê? O que se deseja com isso? Por que um psicólogo que vai dedicar-se ao seu consultório ou ao departamento de pessoal de uma indústria, por exemplo, deve conhecer umas tantas noções de Estatística? Admitindo mesmo que algo eles devam conhecer, o que ensinar? Seriam esses mais que maltratados testes de hipóteses ou uma muito incompreendida análise de variância, ou deveríamos pô-los a par da importância científica e técnica e das condições epistemológicas inerentes às aplicações dos métodos estatísticos em suas conjunturas profissionais e científicas? Deveríamos, por outro lado, dar-lhes a prevalência de uma adequada e intuitiva formação estatística no campo exploratório ou descritivo dos dados, ou ensinar-lhes esses arremedos caricatos da ciência inferencial, de uma indução empírica mais do que capenga?

Não seremos tolos ao ponto de achar que esse estado de coisas vá ser alterado rapidamente. Pelo contrário, essa última questão, que bem pode chamar-se de inércia do ensino, serve a muitos propósitos e interesses que não podem ser contrariados por sustentarem o comodismo e a omissão dos professores de Estatística e de métodos de pesquisas psicossociais mais qualificados academicamente. E bastaria, para comprovar esse estado de coisas, perguntar quem são os que ensinam os cursos de Estatística dirigidos aos alunos dos departamentos de Psicologia, de Sociologia, de Ciências Políticas e de Pedagogia das universidades. São eles, os cursos, designados por aqui, no Brasil, de um modo um tanto pejorativo sob a chancela de disciplinas de serviços. Enquanto os seus tutores são, no geral, professores-assistentes e, mais que isso, os menos

experientes, porque os mais antigos rapidamente livram-se do escolho de ensinar o que também não sabem a alunos que ou não querem saber ou não têm condições de aprender. Pode-se dizer, dessa maneira, sem medo de errar ou de cometer-se qualquer tipo de injustiças, que esses cursos são ministrados pelos que não estão aptos a fazê-lo, nem pedagógica e muito menos técnica e cientificamente. E é uma verdadeira purgação, não nos iludamos, ter-se que ensinar essa estatística do equívocos a platéias pouco afeitas às abstrações e desapetrechadas do mínimo conhecimento matemático que o conhecimento da Estatística mais séria requer... O resultado de tudo isso é muito doloroso, porque, se já não são convenientes à formação dos profissionais da área psicossocial os conteúdos programáticos das disciplinas de Estatística, Psicometria e quejandos, agrava-se o problema pela deficiente qualificação científica de seus responsáveis. E, mais urna vez, em razão dos problemas expostos, os professores dos departamentos psicossociais agem do modo mais estranho, porque, ao acusarem de alienantes e inconvenientes os insubstituíveis métodos estatísticos — sem os quais não existiriam as suas ciências —, afastam-se das verdadeiras causas dos males de que muito padecem. Há, desse modo, um conluio de omissões e equívocos que atenta contra a qualidade da formação profissional e que vai refletir-se, de forma mais grave, nos trabalhos de pesquisa de que não podem eximir-se os professores das ciências psicossociais. Assim, uma vez mais, volta-se à literatura, porque não são as uvas que estão verdes e nem só a raposa foi incapaz de alcançá-las, mas, estranhamente, tampouco o vinicultor pôde usufruir os múltiplos benefícios de seu cultivo...

Sou de opinião que os problemas expostos só serão superados mediante um esforço conjunto de estatísticos e cientistas psicossociais para aproximarem mais as suas ciências, ou melhor, para difundirem-nas bem mais aprofundadamente, bem mais criticamente... Cabe, sobretudo aos estatísticos, um grande esforço de absorção dos novos conhecimentos da difusa estatística observacional e dos correspondentes métodos de análise. Isto significa renunciar — no que diz respeito ao tratamento das ciências psicossociais — às já clássicas (e infladas de sucesso) posturas estatísticas desenvolvidas sob a ótica das ciências experimentais. Dos cientistas psicossociais requerem-se, de outra parte, uma ampla mudança de mentalidade bem como a coragem para reformular, e de modo radical, os conteúdos programáticos de todas as disciplinas de caráter estatístico de seus cursos. Isso, entretanto, é tema para a próxima seção...

O QUE SE DEVE ENSINAR

Antes de quaisquer outras considerações referentes aos conteúdos programáticos que devem constituir os cursos de Estatística para a formação dos profissionais da área psicossocial, cumpre fazer alguns

comentários pertinentes ao modo como a Estatística interage com essas ciências. Com freqüência, nesses domínios, cursos e livros toam nomes diversos que podem ser classificados nas três categorias seguintes, onde a palavra Psicossociologia quer significar, como sempre, cada uma das respectivas ciências do comportamento de base psicológica antes mencionadas: Estatística Aplicada à Psicossociologia, Estatística Psicossociológica e Psicossociometria ou Estatística Psicossociológica.

Na realidade, esses títulos são objeto de muita confusão e pouco entendimento. Trata-se, na verdade, de três categorias essencialmente diferentes de conteúdos estatísticos, todos eles essenciais a um completo entendimento do papel que a Estatística exerce nas ciências psicossociais, além de complementares entre si. Com efeito, por Estatística Psicossociológica entende-se o uso dos métodos estatísticos para gerar os dados ou observações específicas dessas ciências; por Psicossociometria ou Psicossociologia Estatística quer-se referir à formulação de modelos e teorias psicossociológicas, de cunho estatístico e probabilistico; e, finalmente, por Estatística Aplicada à Psicossociologia traduz-se a aplicação dos instrumentos estatísticos de natureza exploratória e inferencial aos dados gerados pela Estatística Psicossociológica.

Desse modo, por exemplo, a Análise de Variâncias é um método próprio de análise estatística e, em razão disso, pode ser aplicada a certos contextos observacionais da Psicologia. Nessas circunstâncias, a Análise de Variância é um método de estatística aplicada à Psicologia. Em contraposição a isso, por outro lado, a Teoria da Confiabilidade dos testes subjetivos é Psicossociologia Estatística, por ser uma teoria de cunho psicossociológico, tendo caráter estatístico.

Essa distinção não é arbitrária e já é tradicionalmente usada na Economia, onde se conhecem e não se misturam as três correspondentes categorias estatísticas. Físicos e astrônomos, de igual maneira, já há muito tempo lidam com as suas particulares Fisicometria e Astronometria, só que sob as denominações alternativas e menos novidadosas de Física Estatística e Astronomia Estatística. Foram os economistas os responsáveis por essa onda de metrias que roubou à Estatística o seu claro e essencial sentido de aplicabilidade. Feitas essas considerações, já se pode lembrar que existe um pensar estatístico inerente a cada ciência. Essa forma de ver os fenômenos particulares de um domínio é o resultado do reconhecimento da diferenciação entre indivíduos ou objetos de investigação constituintes do que se pode denominar coletivo e no concernente às suas propriedades ou características. Sendo assim, cada indivíduo ou objeto possui uma medida de cada característica geral, e a consideração do conjunto dessas medidas leva à noção de distribuição de freqüências das correspondentes propriedades. Infelizmente, no geral, um coletivo não pode ser abordado em sua integralidade e, desse modo, a ciência deve fazer uso de suas partes abordáveis, alcançáveis ou disponíveis, isto é, ela

examina aquilo que tecnicamente se designa por amostras. Decorrem, dessas, duas ordens gerais de conseqüências. Na primeira, uma amostra é investigada em suas propriedades em si, isto é, ela é explorada ou descrita desde o ponto de vista de sua consideração como um coletivo restrito. Nesse caso, desenvolve-se o que se denomina de Estatística Exploratória ou Estatística Descritiva, ou, ainda, Análise Estatística de Dados. Por ela busca-se intuir as propriedades que vão construir as hipóteses. No segundo modo de ver uma amostra, quer-se induzir as propriedades do coletivo a partir das revelações originadas na exploração, isto é, trata-se de criar um método de indução empírica chamado de Inferência Estatística. Segundo essa forma de ver, uma amostra deve ser obtida por um mecanismo de aleatorização. Esta é a grande contribuição da Estatística para resolver o clássico problema científico da indução empírica, que, desde as formulações de David Hume, John Stuart Mill, Karl Pearson, Rudolf Carnap, Karl Popper e outros eminentes filósofos da ciência, ainda não se encontra satisfatoriamente resolvido. A dificuldade enfrentada pelas ciências psicossociais, no concernente à Inferência Estatística, reside, de um modo bem abrangente, no fato de que as suas amostras não são probabilizadas, ou melhor, não são conhecidas as suas probabilidades de seleção. É por essa razão primordial — apesar de existirem outras já destacadas nas seções anteriores — que tenho definido a prevalência do estudo estatístico de caráter exploratório no ensino universitário das ciências psicossociais. Pelo menos, essa deve ser a posição prevalente em cursos profissionais. Apoiando-se, assim, nessas conclusões, podem ser enunciadas as linhas gerais que devem nortear, sobretudo, a formação estatística dos cientistas psicossociais. Pode-se, desse modo, oferecer às considerações mais justas de quantos se interessam por tais problemas as seguintes diretrizes gerais sobre o ensino de Estatística e Psicometria:

i) a formação estatística deve ser eminentemente exploratoria;

ii) devem ser privilegiados os estudos referentes à construção dos instrumentos conceituais de mensuração;

iii) devem ser bem compreendidas as limitações dos instrumentos de mensuração e a natureza dos erros de mensuração;

iv) devem ser destacados os métodos de exploração das variáveis expressáveis nas escalas nominal e ordinal.

Em face desses princípios, o leitor já tem como aquilatar o divórcio definitivo entre o que ele aprendeu ou vem ensinando e aquilo que é o mais adequado às suas necessidades científicas e profissionais. Deve-se ter coragem para renunciar ao que vem sendo oferecido como conhecimento estatístico sob as variadas denominações que têm camuflado o imenso fosso entre a necessidade e a oferta desse conhecimento impres-

cindível à ciência, mas que não nos iludamos quanto a mudanças radicais. Elas não ocorrerão, até mesmo porque os departamentos de Estatística não contam com especialistas capazes de levar a bom termo essa hercúlea tarefa.

Antes que novos elementos sobre este assunto possam ser acrescidos, convém abordar a importante questão do livro-texto e das imprescindíveis leituras de artigos técnico-científicos nas revistas especializadas. Os livros disponíveis não tratam de temas como a Escalagem Multivariada, a Teoria das Respostas aos itens, a Análise de Estruturas Latentes, os Modelos Lineares Multiequacionais, os Modelos Loglineares, a Análise de Correspondências Múltiplas, e outras sofisticadas técnicas de cujo uso científico ou aplicado muito se beneficiariam as ciências psicossociais. Não se pode ser injusto, entretanto, sobretudo com autores como Raymond Boudon, Hubert Blalock Jr., Herbert Solomon, Louis Guttman, Paul Lazarsfeld, WarrenTorgerson, Johan Galtung, C.H. Coombs, A. Birnbaum, L.L. Crombach, K.G. Jereskog, F.M. Lord, M.R. Novick, G. Rash, Louis Thurstone, R. Thorndike, J. Guilford e H. Gulliksen. Esses autores, se bem que profundos em suas exposições, ora pecam pela extrema especialização em alguns poucos tópicos, ora por transformarem em matemática ou estatística complexa aquilo que poderia ser mais simplificado sem perda do rigor lógico, ora por abordar insuficientemente temas tão importantes. Nesse rol amplo, estão incluídos quase todos os principais teóricos do domínio da Psicometria e da Estatística Social.

Essas digressões levam-nos à conclusão de que, no concernente a livros-textos, há uma total impertinencia de seus conteúdos estatísticos relativamente às idéias aqui expostas. Em particular, deve-se ressaltar, são desatualizados e insuficientes. No que se refere, por outro lado, a artigos nas revistas científicas, observa-se, nas mais importantes, um sofisticado e, muitas vezes, estéril tratamento estatístico-matemático que afugenta de suas leituras o leitor de formação humanística. Cumpre dizer que, algumas vezes, também, os artigos tratam de generalizações absolutamente desnecessárias ou, em certas situações, conferem a temas de fácil compreensão um tratamento estatístico-matemático que verdadeiramente deforma as suas concepções básicas. Este é, entretanto, um quadro geral de todas as revistas científicas que atendem às pressões acadêmicas para que se publiquem artigos originais sem importarem-se com o leitor médio ou com as idéias que deveriam ser prevalecentes, em detrimento dos formalismos matemáticos. Elas são, desse modo, uma espécie comum de jogo de compadres, onde uns poucos se entendem e muitos, para não ficarem atrás, além de pagarem, fingem atualizarem-se. Assim, urge que esses especialistas promovam os seus melhores esforços em prol da publicação de livros modernos de Estatística Psicossocial e que contemplem todas as ciências por ela envolvida, numa postura mais nomotética do que ideográfica, na linha do que eu tenho entendido como a Psicossociologia.

Com efeito, é desse modo que se unem umbilicalmente, pela base da subjetividade mental, grande parte das teorias sociológicas e psicológicas, e as correspondentes mensurações, nesses domínios, exibem bases científicas e metodológicas equivalentes e muitas dificuldades comuns a serem superadas mediante a imprescindível união de esforços. Posta dessa maneira, a visão psicométrica estrita, de tradição psicológica, de cunho ideográfico, cede lugar, neste momento, à expansão larga de seus limites, tornando à Psicometria velha apenas a semente de uma Psicossociometria, cidadã nomotética das ciências do comportamento de bases psicológicas.

CAPÍTULO 7

TEORIA DA RESPOSTA AO ITEM — IRT: UMA INTRODUÇÃO

Luiz Pasquali
Instituto de Psicologia
Universidade de Brasília

A IRT (Item Response Theory) já tem uma longa história. Ela iniciou com os trabalhos de Lord (1952, 1953) nos Estados Unidos e Rasch (1960) na Holanda, que a utilizaram para testes de desempenho e de aptidão. Contudo, apenas ultimamente, a partir de meados dos anos 80, a IRT vem se tornando a técnica predominante no campo dos testes. A razão da demora desta teoria em ser amplamente utilizada em psicometria consiste na enorme complexidade de manipulação de seus modelos matemáticos, inviáveis sem os requintados programas de computador, e estes só começaram efetivamente a entrar no mercado nos anos 80.

Atualmente, a IRT parece que veio para ficar e substituir grande parte da teoria clássica da psicometria — isto é um fato que já ocorre no Primeiro Mundo (USA, Canadá, Europa, Japão, Israel, Austrália); no restante do mundo ela é raramente utilizada, e no Brasil (América Latina em geral) ela sequer é conhecida. Este capítulo visa precisamente iniciar no País o conhecimento e, esperamos, o uso desta técnica no campo da psicometria.

As publicações em IRT vêm crescendo e tomando conta das revistas especializadas, como a *Psychometrika.* Há centros importantes de pesquisa nesta área nos USA (University of Massachusetts at Amherst), Holanda e Espanha (Universidade de Oviedo). Existe, inclusive, uma sociedade internacional, a International Test Commission (ITC), que filia seguidores da IRT. De fato, no Congresso Internacional da ITC, em Oxford (Inglaterra), de julho de 1993, havia mais de 120 participantes de cerca de 46 países. Da América Latina só estavam representados o Brasil e a Argentina, com dois participantes cada.

O enorme impacto que a IRT vem tendo em psicometria se deve ao fato de ela superar certas limitações teóricas graves que a psicometria tradicional contém. Hambleton, Swaminathan e Rogers (1991) salientam especialmente quatro dessas limitações:

1) Os parâmetros clássicos dos ¡tens (dificuldade e discriminação) dependem diretamente da amostra de sujeitos utilizada para estabelecê-los *(group-dependenf).* Daí, se a amostra não for rigorosamente representativa da população, aqueles parâmetros dos itens não podem ser considerados válidos para esta população. Como conseguir amostras representativas é um problema prático grave para os construtores de testes; a dependência dos parâmetros dos itens na amostra obtida se torna um empecilho de grandes proporções para a elaboração de instrumentos psicométricos não enviesados.

2) A avaliação das aptidões dos testandos também depende do teste utilizado *(test-dependent).* Assim, testes diferentes que medem a mesma aptidão irão produzir escores diferentes da mesma aptidão para sujeitos idênticos. Testes com índices de dificuldade diferentes evidentemente produzirão escores diferentes. Mesmo no caso das formas paralelas, há sempre a dificuldade de que o montante de erros nas duas formas dificilmente seja o mesmo, o que produzirá novamente escores diferentes.

3) A definição do conceito de fidedignidade ou precisão na teoria clássica dos testes constitui também uma fonte de dificuldades. Ela é concebida como a correlação entre escores obtidos de formas paralelas de um teste ou, mais genericamente, como o oposto do erro de medida. Ambos os conceitos apresentam dificuldades. Primeiramente, é praticamente impossível satisfazer as condições de definição de formas paralelas e, no caso do erro de medida, é postulado que este seja idêntico em todos os examinandos — postulado improvável (Lord, 1984).

4) Outro problema da teoria clássica dos testes consiste em que ela é orientada para o teste total e não para o item individual. Toda a informação do item deriva de considerações do teste geral, não se podendo assim determinar como o examinando se comportaria diante de cada item individual.

Estas e outras dificuldades dos modelos e técnicas clássicos de medida incitaram os psicometristas à procura de teorias alternativas que pudessem permitir estabelecer (Hambleton et ai., 1991):

a) características do item sem ser dependente da amostra de sujeitos utilizados;
b) escores dos examinandos independentes do teste utilizado;
c) um modelo em nível do item em vez do teste;
d) um modelo que não exija formas rigorosamente paralelas para avaliar a fidedignidade;
e) um modelo que ofereça uma medida de precisão para cada escore de aptidão.

Essas características são precisamente oferecidas pela Teoria da Resposta ao Item (Hambleton, 1983; Hambleton, Swaminathan, 1985; Lord, 1980; Wright, Stone, 1979; Hambleton, Swaminathan, Rogers, 1991; Muñiz, 1990).

CARACTERÍSTICAS DA IRT

Teoria da IRT

Contrariamente à teoria clássica de psicometria, a IRT trabalha com traços latentes e coloca dois axiomas fundamentais: 1) o desempenho do sujeito numa tarefa (item do teste) se explica em função de um conjunto de fatores ou traços latentes (aptidões, habilidades, etc.) — o desempenho é o efeito, e a causa são os traços latentes; 2) A relação entre o desempenho na tarefa e o conjunto dos traços latentes pode ser descrita por uma equação monotônica crescente, chamada de ICC (Item Characteristic Function ou Item Characteristic Curve — a curva característica do item) e exemplificada na Figura 7-1, onde se observa que sujeitos com aptidão maior terão maior probabilidade de responder corretamente ao item e vice-versa (Θ_i é a aptidão e $P_i(\Theta)$ a probabilidade de resposta correta).

Hã um número ilimitado de modelos para expressar esta relação, dependendo do tipo de função matemática utilizada e/ou do número de parâmetros que se quer descobrir para o item.
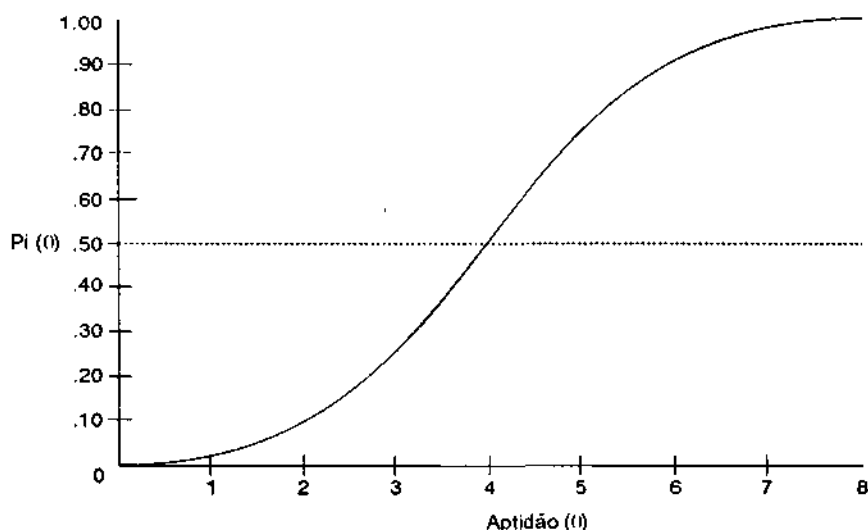
Fig. 7-1 —A curva ICC.

Uma preciosa vantagem da IRT sobre a teoria clássica consiste em que os modelos utilizados permitem desconfirmação. Na verdade, a demonstração da adequação do modelo aos dados *(goodness-of-fif)* é um passo necessário nos procedimentos desta teoria.

Propriedades da IRT

Entre as características da IRT, duas são de especial relevância: unidimensionalidade e independência local.

Unidimensionalidade

A IRT postula que há apenas uma aptidão responsável pela realização de um conjunto de tarefas (itens). Parece pacífico que qualquer desempenho humano é sempre multideterminado ou multimotivado, dado que mais de um traço latente entra na execução de qualquer tarefa. Contudo, para satisfazer o postulado da unidimensionalidade, é suficiente admitir que haja uma aptidão **dominante** (um fator dominante) responsável pelo conjunto de itens. Este fator é o que se supõe estar sendo medido pelo teste. O postulado da unidimensionalidade ainda continua importante, dado que a IRT, embora estudos estejam sendo feitos nesta área, ainda não possui soluções adequadas para modelos multidimensional.

Independência local

Este postulado afirma que, mantidas constantes as aptidões que afetam o teste, as respostas dos sujeitos a quaisquer dois itens são estatisticamente independentes. Seja 0 o conjunto de aptidões que afetam um conjunto de itens, Uj a resposta de um sujeito ao item **i** (i = 1, 2, ..., n) e $P(U_i|0)$ a probabilidade de resposta do sujeito **i** com
aptidão 0. $P(U_i = 1|0)$significa a probabilidade de uma resposta correta e
$P(U_i = 0|0)$ a probabilidade de uma resposta errada. A independência local pode ser matematicamente afirmada como

$$Prob(U_1, U_2, ..., U_n|\Theta) = \\ = P(U_1|\Theta)P(U_2|\Theta)...P(U_n|\Theta) = \\ = \prod_{i=1}^{n} P(U_i|\Theta)$$

A independência local significa que, para examinandos com uma aptidão dada, a probabilidade de resposta a um conjunto de itens é igual aos produtos das probabilidades das respostas do examinando a cada item individual. Assim, se um sujeito acertou os itens 1 e 2 e errou o 3, a configuração de suas respostas é U, = 1, U_2 = 1, U_3 = 0 (1, 1, 0), e a independência local implica que

$$P(U_1=1,\ U_2=1\ U_3=0\,|\,\theta) =$$
$$= P(U_1=1\,|\,\theta)\,P(U_2=1\,|\,\theta)\,P(U_3=0\,|\,\theta) =$$
$$= P_1P_2Q_3\ ,$$

sendo $P(U_i=1\,|\,\theta)$ e $Q_i = 1 - P_i$.

Embora pareça improvável que os comportamentos de um mesmo sujeito não estejam correlacionados, a independência local afirma que, se houver correlação, esta se deve à influência de fatores outros que não o fator dominante. Se estes outros fatores forem controlados (mantidos constantes), o fator dominante será a única fonte de variação, e as respostas se tornam independentes. Assim, a independência local implica a unidimensionalidade (Lord, 1980; Lord, Novick, 1968)

MODELOS DA IRT

Embora seja ilimitado o número de modelos matemáticos que podem expressar a relação de probabilidade de sucesso em um item e a aptidão medida pelo teste (isto é, a ICC), na prática há três que predominam. Estes se distinguem pelo número de parâmetros que utilizam para descrever o item — os modelos logísticos de I, 2 e 3 parâmetros, a saber: a) que avaliam do item somente a dificuldade, b) a dificuldade e a discriminação, c) ou a dificuldade, a discriminação e a resposta correta dada ao acaso.

Modelo logístico de 1 parâmetro

Este modelo, inicialmente criado por Rasch (1960) e expresso como modelo de ogiva, foi descrito para um modelo logístico por Wright (1977a, 1977b) e permite tratamento matemático mais fácil. Sua fórmula é

$$P_i(\theta) = \frac{e^{D(\theta - b_i)}}{1 + e^{D(\theta - b_i)}} \qquad (I = 1.\ 2........n),$$

onde $P_i(\Theta)$ é a probabilidade de um examinando com aptidão 0 responder o item I e é representado como uma curva tipo S; b, é o parâmetro de dificuldade do item I; n é o número de ¡tens no teste; **e** é um número transcendental com valor de 2,718; D, constante que vale 1,7.

$P_i(\Theta)$ produz uma curva chamada curva característica do item (ICC — Item Characteristic Curve), conforme a Figura 7-1.0 parâmetro b¡ do item corresponde ao ponto, na escala de aptidão 0, onde a probabilidade de resposta é 0,50. Quanto maior for o b_i, maior deve ser o nível de aptidão exigido para que o examinando tenha a chance de 50% de acertar o ¡tem. Transformando a escala da aptidão em escores padrões, com média = 0 e desvio padrão = 1, os valores de b, tipicamente se situam entre -2 (itens

fáceis) e +2 (¡tens difíceis), conforme Figura 7-2, onde o item 1 exige aptidão de cerca de -1,10 e o item 2 aptidão 0,50, sendo este item mais difícil que o item 1.
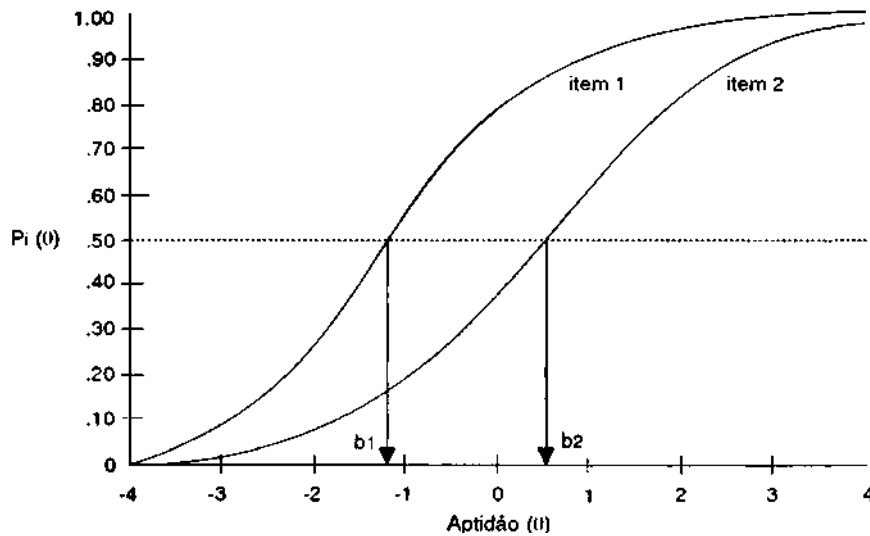
Fig. 7-2 — Parâmetro de dificuldade (b) de dois itens.

A constante D foi incluída na fórmula para tornar a curva logística igual à curva normal acumulada (ogiva) utilizada nos estudos pioneiros da IRT.

Modelo logístico de 2 parâmetros

Birnbaum (1968) desenvolveu a equação que serve para avaliar dois parâmetros do item: dificuldade e discriminação. A fórmula é

$$P_i(\theta) = \frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}},$$

onde a, é o parâmetro de discriminação do item, que pode variar de -∞ a +∞, mas tipicamente varia entre 0 e 2. Valores negativos indicariam que a probabilidade de acertar um item estaria inversamente relacionada com a aptidão, o que soa estranho. A Figura 7-3 mostra os dois parâmetros do item (b_i e a_i). O a_i é representado pela inclinação da curva no ponto de inflexão, onde a probabilidade de resposta correta é 0,50.

Na ilustração, o item 2 é mais difícil que o item 1 (parâmetro b_i, mas muito menos discriminativo (parâmetro *a),* pois a inclinação da sua curva é bem menor que a do item 1.
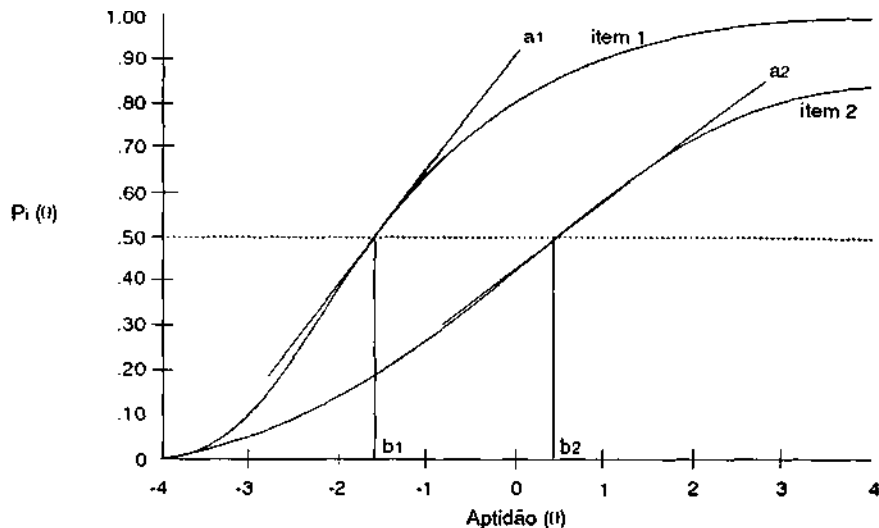
Fig. 7-3 — Parâmetros de dificuldade (b) e discriminação (a) de dois itens.

Modelo logístico de 3 parâmetros

Desenvolvida por Lord (1980), a fórmula deste modelo é:

$$P_i(\theta) = c_i + (1 - c_i)\frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}},$$

onde c_i é o parâmetro do item que avalia a resposta correta dada ao item por acaso e é expresso pela assíntota inferior da curva. Se esta assíntota cortar a ordenada acima do ponto 0, há presença de acertos ao acaso (Figura 7-4). No caso do item 2 na Figura 7-4, há 20% de probabilidade de que o item seja acertado por acaso, sendo esta probabilidade de 0 para os outros dois itens.

DETERMINAÇÃO DOS PARÂMETROS DE ITENS E APTIDÕES

A determinação destes parâmetros constitui apenas uma das etapas na elaboração de instrumentos psicológicos. As etapas da elaboração de instrumentos dentro da IRT dividem-se em três níveis de procedimentos:

1) Procedimentos teóricos, onde se incluem as etapas de a) estabelecimento do sistema ou variável (traço latente) a ser medido;
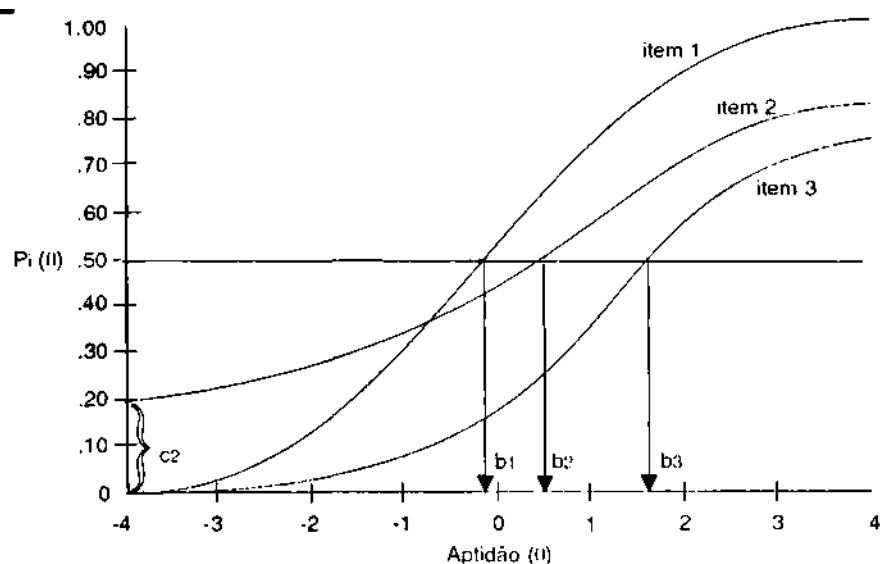
Fig. 7-4 — ICC do modelo de três parâmetros para três itens.

b) desenvolvimento da teoria psicológica sobre este traço;
c) operacionalização do traço através da elaboração dos comportamentos que o representam (elaboração dos itens); e
d) análise teórica dos itens.

2) Procedimentos empíricos, que consistem em
   a) definição da amostra de sujeitos para a coleta da informação sobre o teste que se quer utilizar no futuro na população; e
   b) aplicação dos itens a esta amostra.

3) Procedimentos analíticos, que consistem em
   a) escolha do modelo de IRT;
   b) estabelecimento da dimensionalidade do traço (unidimensionalidade dos itens);
   c) avaliação dos parâmetros dos itens e da aptidão do sujeito (o traço 0);e
   d) demonstração da adequação do modelo aos dados empíricos.

Na IRT, o desempenho do sujeito numa tarefa (item), isto é, a probabilidade de resposta correta [Pj(0)], depende de 1) aptidão do sujeito (0) e 2) dos parâmetros dos itens (a_i, b_i e c_i). Daí, a primeira tarefa do IRT é viabilizar modelos que possam permitir a descoberta dos parâmetros dos itens.

Com base nos dados empíricos, isto é, as respostas da amostra de sujeitos aos itens, se faz a estimativa dos parâmetros destes itens. Isto

consiste em se escolher corno parâmetros para os ¡tens aqueles valores que maximizam a probabilidade de ocorrência dos dados que de fato apareceram nas respostas dos sujeitos. Por exemplo, se se lançar 100 vêzes uma moeda (cara ou coroa) e aparecer 60 caras e 40 coroas, a probabilidade mais verossímil de que apareça cara é de 60/100, isto é, 0,60. Assim, estima-se que 0,60 é o valor mais provável de aparecer cara. Este método de avaliação se chama de máxima verossimilhança *(maximum likelihood),* porque os valores estimados são os mais verossímeis, plausíveis, com respeito aos dados empíricos obtidos.

A estimação dos parâmetros se faz por aproximações sucessivas (iteração), utilizando-se pacotes estatísticos apropriados, tais como BICAL (Wright et al., 1979), para modelos logísticos de 1 parâmetro, e BILOG (Mislevy, Bock, 1984) e LOGIST (Wingersky, Barton, Lord, 1982), para modelos de 1,2 e 3 parâmetros. Estes pacotes produzem tanto os parâmetros dos itens quanto os valores de 0 dos sujeitos (Baker, 1987; Birnbaum, 1968; Lord, 1980; Swaminathan, 1983; Hambleton, Swaminathan, 1985).

A lógica destes procedimentos será ilustrada brevemente, seguindo Muñiz Fernández (1990). Os valores estimados para os parâmetros pelo método da verossimilhança são os que maximizam a probabilidade de que ocorram aqueles valores (respostas) dados pelos sujeitos. Então procura-se uma função matemática que produza estes máximos, como ilustrado na Figura 7-5 (onde o sujeito 1 tem seu máximo em 0 = - 1,50 e o sujeito 2 em 0=1).
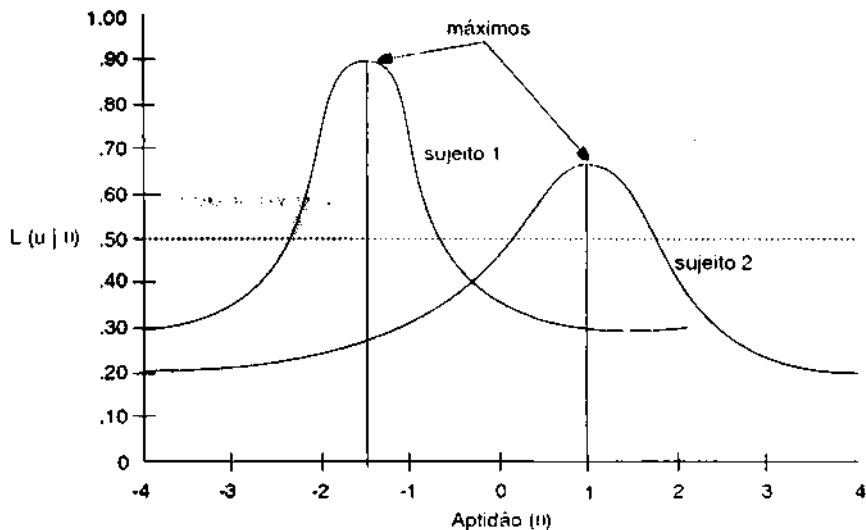
Fig. 7-5 — Máximos da função de verossimelhança.

Seja Uj a resposta ao item. Em testes de aptidão, Uj = 1 se o item for corretamente respondido ou U, = 0 se o sujeito errar o item. A equação ICC dá precisamente a probabilidade de acerto e erro para um dado valor de aptidão 0, sendo $P(U_1=1|\theta) = P(\theta)$ e $P(U_1=0|\theta) = 1 - P(\theta) = Q(\theta)$, onde se entende $P(U_1=1|\theta)$ como a probabilidade de uma resposta correta para um tal valor de G. Exemplo: a probabilidade de acertar o item na Figura 7-6 é de 0,60 e a de errar é de 0,40 para um valor 0 de 1:

$$P(U_1=1|\theta=1) = 0{,}60$$
$$P(U_1=0|\theta=1) = 0{,}40$$
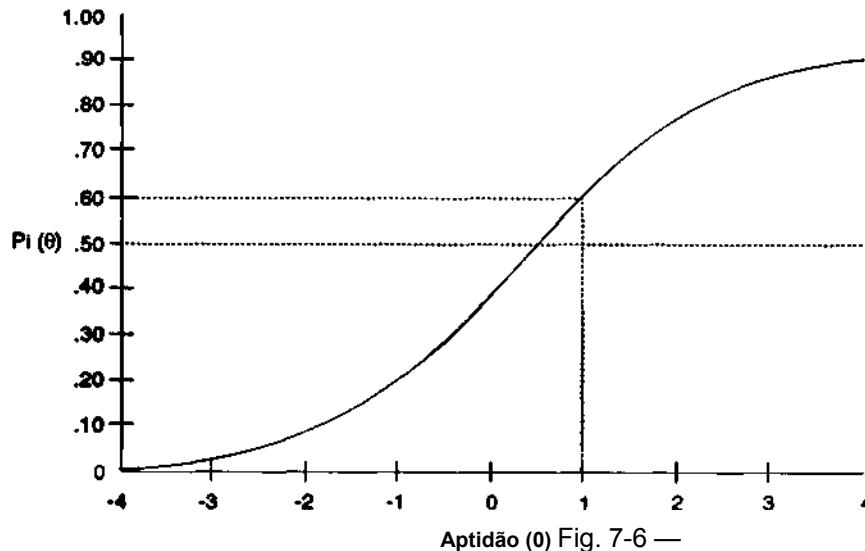
Fig. 7-6 — Probabilidade 0.60 de acerto do item a 0 = 1.

Como a resposta a um item para um dado valor de 0 é uma prova de Bernoulli, segue que

$$P(U_i|\theta) = P(U_i=1|\theta)^{U_i} P(U_i=0|\theta)^{(1-U_i)} =$$
$$= [P(\theta)]^{U_i}[Q(\theta)]^{(1-U_i)}$$

No nosso caso,

$$P(U_1=1|\theta=1) = (0{,}60)^1(0{,}40)^{(1-1)} = 0{,}60$$
$$P(U_1=0|\theta=1) = (0{,}60)^0(0{,}40)^{(1-0)} = 0{,}40$$

Como um teste tem n itens, a probabilidade de um padrão de resposta é dado pelo produto das probabilidades de cada item (dado o

axioma da independência local). Assim, um padrão 11010 para cinco itens será:

$$P_1(\Theta)P_2(\Theta)Q_3(\Theta)P_4(\Theta)Q_5(\Theta) \quad \text{ou, em geral,}$$

$$L(u_1, u_2, ..., u_n|\Theta) = \prod_{i=1}^{n} [P_i(\Theta)]^{u_i}[Q_i(\Theta)]^{(1-u_i)},$$

que é a função de verossimilhança *(maximum likelihood function).*

Esta equação vem normalmente expressa em termos de logaritmos por ser mais fácil de operar, pois

$$\ln xy = \ln x + \ln y \quad \text{e}$$
$$\ln x^a = a \ln x$$

Assim, a equação de verossimilhança se escreve

$$\ln L(u|\Theta) = \sum_{i=1}^{n} [u_i \ln P_i(\Theta) + (1 - u_i) \ln Q_i(\Theta)],$$

AJUSTE DO MODELO *(model-data goodness-of-fif)*

Estabelecidos os parâmetros do modelo, é preciso demonstrar que o modelo IRT escolhido se adapta aos dados empíricos, isto é, se os valores P(0) estimados pelo modelo não diferem dos valores obtidos empiricamente (a saber, a proporção de sujeitos que de fato acertaram o item).

Para tal demonstração há uma série de procedimentos estatísticos que constituem ainda o ponto fraco da IRT (Muñiz, 1990; Hambleton et al, 1991). Entre eles, há o x^2 e a análise dos resíduos.

O χ^2

Wright e Panchapakesan (1969, apud Muñiz, 1990) utilizam uma estatística parecida com o x^2 para verificar o ajuste do modelo:

$$\chi^2 = \sum_{j=1}^{k} \frac{n_j[P(\theta_j) - Pe(\theta_j)]^2}{[P(\theta_j)][1 - P(\theta_j)]},$$

onde k = número de categorias em que se dividiu 0 Θ,n_i = número de sujeitos dentro da categoria, P(Θj) = valor da ICC para a categoria j dado pela fórmula do modelo, Pe(Θj) = proporção de sujeitos que de fato acertaram o item para a categoria J e x^2 se distribui com k-1 graus de liberdade.

onde In = logaritmo natural e *u* = vetor das respostas.

Exemplo (Muñiz, 1990, p.51-53): Mil sujeitos responderam 20 itens. O programa LOGIST estimou os parâmetros, mostrando que o modelo de dois parâmetros seria o aconselhável. Para o item 10, o programa deu que a=l e b=2. O Θ foi dividido em 5 categorias (usa-se o ponto médio das categorias para os cálculos) e os resultados foram os da tabela 7-1.

Tabela 7-1 — Proporção de 1.000 sujeitos acertando item 10 por categoria de 0

Θ	ny	Pe(Θy)	P(Θy)
4-5	70	0,97	0,99
3-4	90	0,95	0,92
2-3	200	0,70	0,70
1-2	300	0,35	0,30
0-1	340	0,10	0,07
	1.000		

Os valores P(0) calculados pelo modelo de dois parâmetros [$P\Theta_j$)] foram conseguidos usando a fórmula deste modelo, onde os parâmetros para o item 10 foram: a = 1 e b = 2.

Aplicando-se a fórmula do x^2 resulta:

$$x^2 = \frac{340(0,07-0,10)^2}{(0,07)(1-0,07)} + \frac{300(0,30-0,35)^2}{(0,30)(1-0,30)} + \frac{200(0,70-0,70)^2}{(0,70)(1-0,70)} + \frac{90(0,92-0,95)^2}{(0,92)(1-0,92)} + \frac{70(0,99-0,97)^2}{(0,99)(1-0,99)} =$$

$$= 4,70 + 3,57 + 0,00 + 1,10 + 2,83 = 12,2,$$

que, para graus de liberdade k-1 = 5-1 = 4, a probabilidade de tal x^2 (=12,2) a gl = 4 ocorrer por acaso se situa entre 0,02 e 0,01. Portanto, somente no nível de 98% de confiança pode-se afirmar que o modelo de dois parâmetros se ajusta aos dados empíricos do item 10. Pela Figura 7-7, vê-se que os valores empíricos, Pe(0), e os calculados, P(0), são bastante similares.

Análise dos resíduos

Esta análise consiste em verificar se a diferença entre o desempenho real dos sujeitos num item e o desempenho predito pelo modelo não é estatisticamente diferente de 0:

$$r_{ij} = P_{ij} - E(P_{ij}),$$
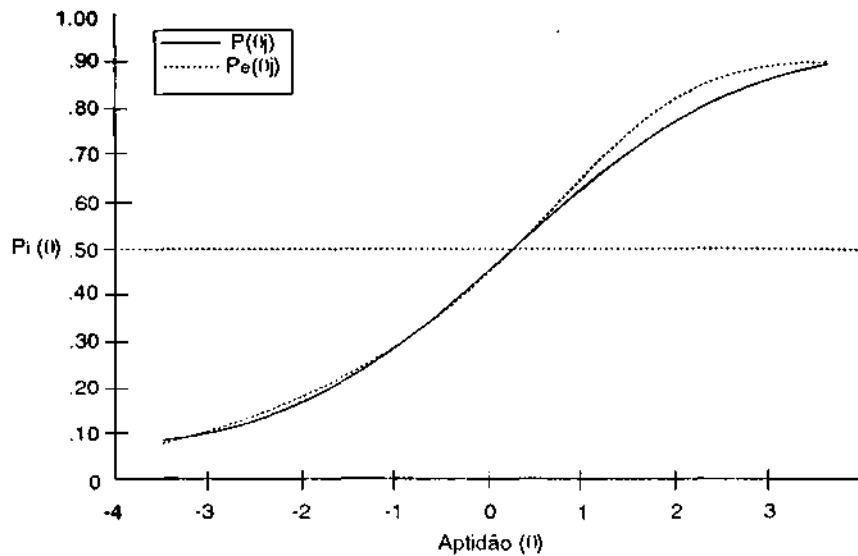
Fig. 7-7 — Valores preditos pelo modelo Pe(Θj) e valores empíricos P(Θj).

onde r, = resíduo, Pij = desempenho real, isto é, proporção de respostas corretas ao item i na categoria j da aptidão Θ, e E(Pij) = desempenho predito pelo modelo.

Normalmente, este resíduo é expresso em dados padronizados (resíduo padronizado - z_{ij}):

$$z_{ij} = \frac{P_{ij} - E(P_{ij})}{\sqrt{E(P_{ij})[1 - E(P_{ij})]/N_j}},$$

sendo N_j o número de sujeitos na categoria. Esta categoria se refere a que o 0 deve ser dividido em categorias (10 a 15), como no caso do x^2.

A Figura 7-8 mostra que os dados empíricos não se coadunam com os preditos, pois as duas linhas não coincidem.

INVARIANZA DOS PARÂMETROS

A invariância dos parâmetros constitui o ponto central da IRT e afirma que se pode estimar 1) os escores dos sujeitos independentemente do teste utilizado e 2) os parâmetros dos itens independentemente da amostra de sujeitos utilizada. Se o modelo IRT utilizado se ajusta aos dados empíricos, então são salvos estes objetivos, ilustrados na Figura 7-9.
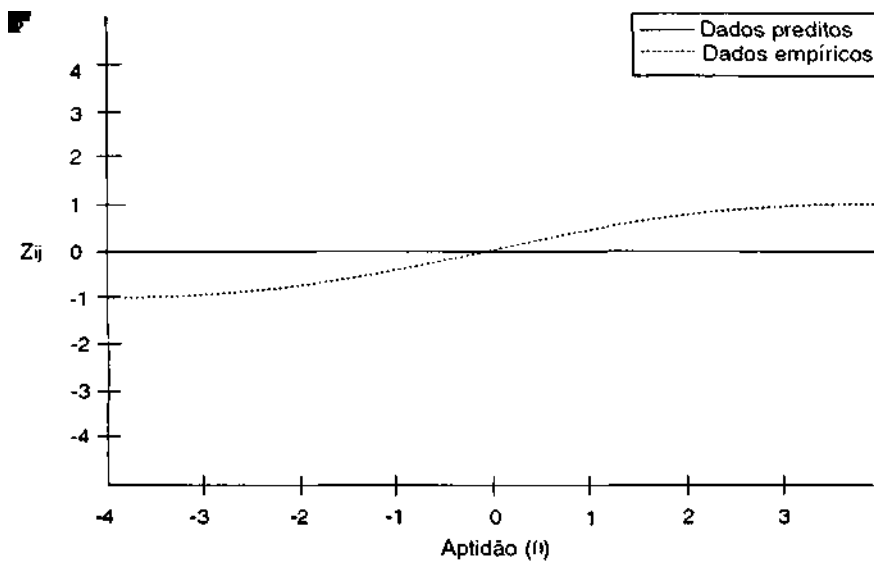
Fig. 7-8 — Discrepância entre modelo e dados empíricos.
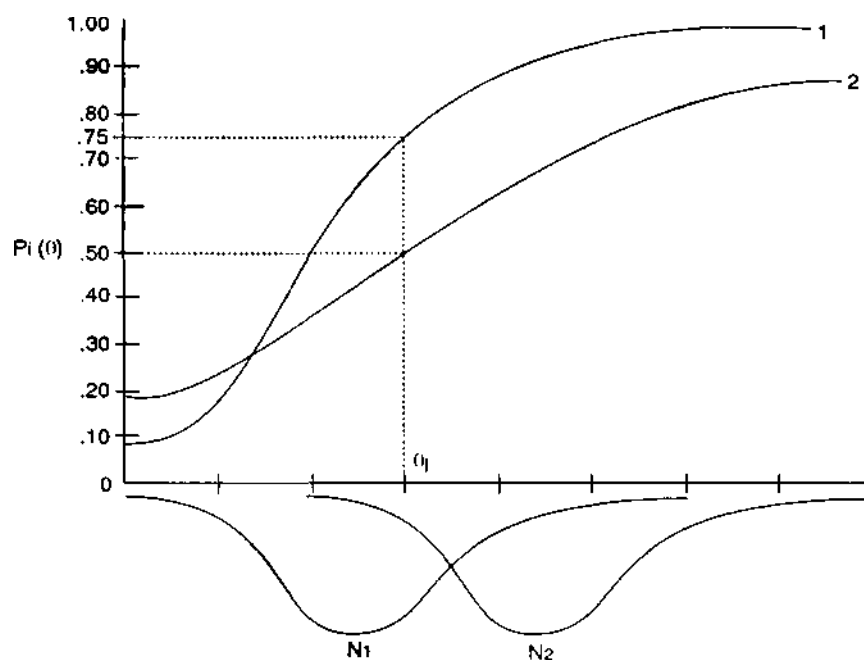
Fig. 7-9 — ICC para dois itens e distribuição de 0
para dois grupos de sujeitos.

A Figura 7-9 mostra que 1) as curvas ICC para os dois itens podem ser obtidas tanto com a amostra N, quanto com a N_2 — conseqüentemente, os parâmetros dos itens independem da amostra utilizada; e 2) o valor da aptidão Θj pode ser obtido utilizando-se tanto o item 1 quanto o item 2 — este Θj corresponde à probabilidade de acerto de 75% do item 1 e 50% do item 2. Continua valendo, contudo, como em qualquer estimação estatística, que quanto maior e mais heterogênea a amostra de sujeitos, mais precisa será a estimação dos parâmetros.

Para demonstrar a invariância da aptidão (0), se aplicam dois testes com itens diferentes, mas que medem a mesma aptidão, a uma mesma amostra de sujeitos, e os resultados mostrarão se há ou não coincidência. Se houver coincidência, então os itens dos dois testes se distribuirão em torno de uma linha reta num sistema de coordenadas, como na Figura 7-10, e uma indicação numérica será dada pela correlação de Pearson entre as duas avaliações.
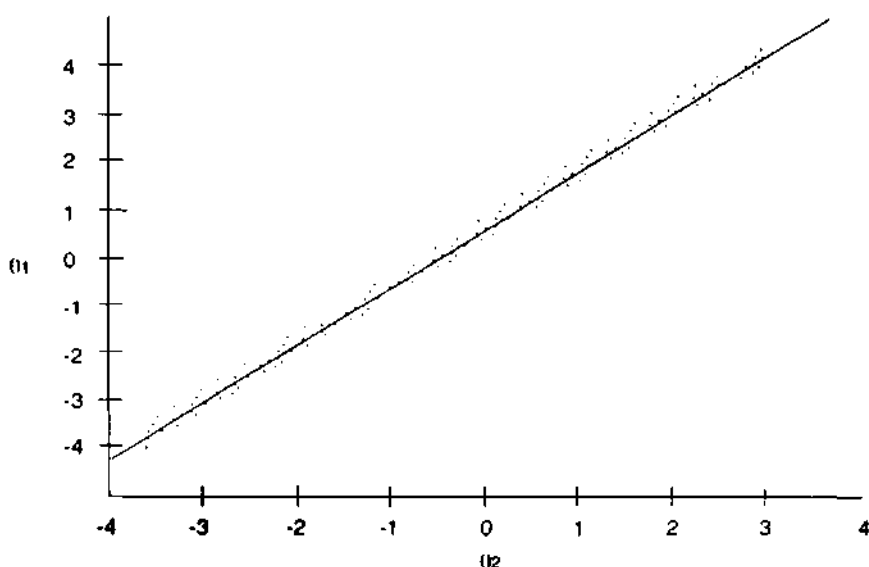
Fig. 7-10 — Valores 9 obtidos por testes diferentes.

Evento similar ocorre na demonstração da invariância dos parâmetros dos itens (a, b, c). Aqui se usam duas amostras de sujeitos para responder ao mesmo teste. Se os parâmetros dos itens (por exemplo, a dificuldade b — vide Figura 7-11) são os mesmos nas duas amostras, novamente surgirá uma linha reta nas coordenadas e a correlação de Pearson estima a coincidência dos parâmetros.
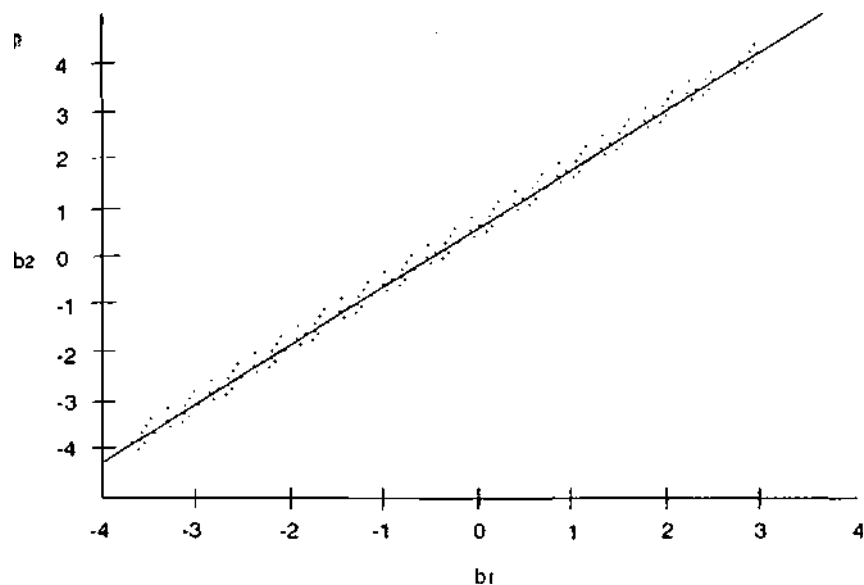
Fig. 7-11 — Valores b¡ obtidos em duas amostras diferentes de sujeitos.

FUNCOES DE INFORMAÇÃO E EFICIENCIA

Um poderoso método para descrever itens e testes bem como para selecionar itens é dado pela função de informação do item e do teste. Ela permite analisar quanto um item ou um teste traz de informação para a medida da aptidão.

Função de informação do item

A fórmula da função é

$$I_i(\theta) = \frac{[P_i'(\theta)]^2}{P_i(\theta)Q_i(\theta)},$$

onde $I_i(\Theta)$ é a "informação" fornecida pelo item **i** no nível da aptidão G, Pi(Θ) é a derivada de Pi(Θ) com relação a Θ, Pi(Θ) é a ICC e $Q_i(\Theta) = 1 - P_i(\Theta)$.

No caso do modelo logístico de três parâmetros, a equação se simplifica para

$$I_i(\theta) = \frac{2{,}89a_i^2(1 - c_i)}{[c_i + e^{1,7a_i(\theta - b_i)}][1 + e^{-1,7a_i(\theta - b_i)}]^2}$$

Esta equação mostra a importância que têm os três parâmetros sobre o montante de informação do item. Na verdade, a informação 1) é maior quando b_i se aproxima de 0; 2) é maior quanto maior for o a_i; e 3) aumenta com a diminuição de c_i, para Θ, como mostra a figura 7-12.
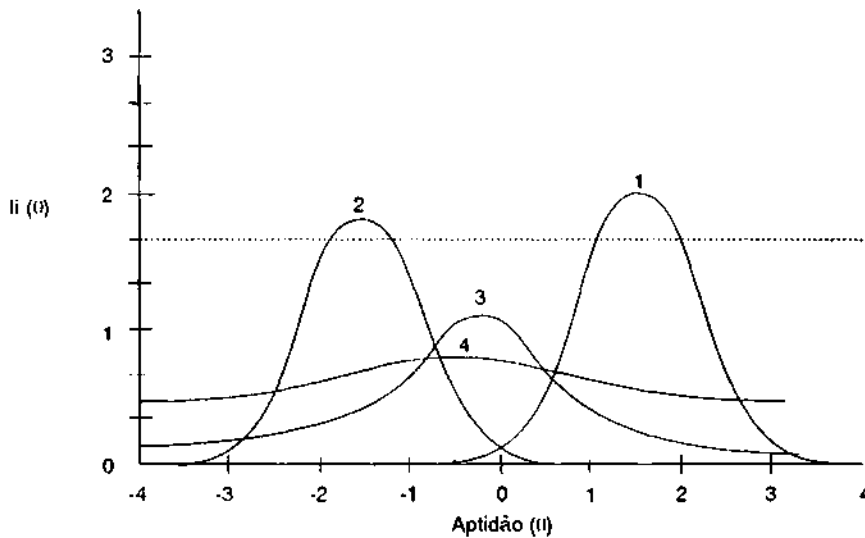
Fig. 7-12 — Função de informação para quatro itens.

Observa-se na Figura 7-12 que o item 1 fornece a maior informação a Θ = 1,5 e o item 2 a Θ = -1,5. O item 3 dá bem menos informação que os itens 1 e 2, porque o c, > 0, mas dá mais informação do que estes itens no nível deΘ = 0. O item 4, devido ao grande c (c_4, = 0,15), tornou-se inútil no teste, pois não produz qualquer informação a mais do que a já produzida pelos itens 1, 2 e 3 em qualquer nível de Θ.

Função de informação do teste

A informação fornecida pelo te $I(\theta) = \sum_{i=1}^{n} I_i(\theta)$ nte a soma das informações fornecidas por cada item c

Outra maneira de representar esta função de informação do teste é através do erro padrão de medida, chamado na IRT de erro padrão de estimação. A 1(Θ), na verdade, é o inverso deste erro:

$$EPE(\theta) = \frac{1}{\sqrt{I(\theta)}}$$ = erro padrão de estimação.

Similarmente ao erro padrão de medida da teoria psicométrica clássica, o EPE permite estabelecer intervalos de confiança em torno dos escores 0 dos sujeitos.

Exemplo: Um teste de 50 itens aplicado a 100 sujeitos deu 1(Θ = 4 para Θ= 3. Para nível de confiança de 95%, qual o intervalo em que cai o 0?

Resposta:

1) para 95%, z= **±1.96**

2) erro máximo = $z_{EPE}(\theta) = z\frac{1}{\sqrt{I(\theta)}} = 1{,}96\frac{1}{\sqrt{4}} = 0{,}98$

3) assim, o intervalo para 0 será 3 ± **0,98,** ¡sto é, 2,02 a 3,98.

Eficiência relativa

A I(0) permite comparar a relativa eficiência de um teste com relação a outro em sua capacidade de estimar a aptidão 0:

$$ER(\theta) = \frac{I_1(\theta)}{I_2(\theta)},$$

onde ER(Θ) = eficiência relativa e I_1,(Θ) e I_2(Θ) são funções de informação dos testes 1 e 2, respectivamente.

Exemplo: Se 1,(Θ) = 30 e I_2(Θ) = 15, então ER(Θ) = 30/15 = 2. O teste 1 é duas vezes mais eficiente que o teste 2 para estimar 0, isto é, o teste
1 poderia ser reduzido pela metade ou o teste 2 aumentado 50% para ambos terem o mesmo nível de eficiência.

TRANSFORMAÇÕES DO 0

O valor 0 de um sujeito define a probabilidade de ele acertar um dado item. A escala 0, onde se situa o escore do sujeito, veio expressa neste capítulo em termos de escores padrões com média = 0 e desvio padrão = 1. Entretanto, para facilitar o uso prático do 0, este pode ser expresso em termos de escalas mais apropriadas e mais inteligíveis, fazendo uso de transformações lineares ou não-lineares.

As transformações lineares consistem em expressar o 0 acrescentando-se a ele uma constante e, além de uma constante, também um fator multiplicativo. Contudo, ao se fazer tal operação no 0, é necessário que se faça o mesmo aos parâmetros dos itens para, assim, manter a invariância da escala 0.

Assim, tem-se,

Para modelo de 1 parâmetro:

$$\theta' = m(\theta) + k$$
$$b' = m(b) + k$$

Para modelo de 2 parâmetros:

$$\theta' = m(\theta) + k$$
$$b' = m(b) + k$$
$$a' = a/m$$

Para modelo de 3 parâmetros:

$$\theta' = m(\theta) + k$$
$$b' = m(b) + k$$
$$a' = a/m$$
$$c' = c$$

Por exemplo, Woodcock (1978) fez a seguinte transformação para seu teste:

$$w = 20 \log (e^{\theta}) + 500 =$$
$$w_b = 9{,}1\theta + 500.$$

Entretanto, a transformação linear mais interessante é aquela que transforma o 0 no escore verdadeiro *(x)* da psicometria clássica:

$$\tau = \sum_{j=1}^{n} P_j(\theta) \ .$$

Contudo, enquanto o 0 se estende de -∞ a +∞, o τ se situa entre 0 e n (número de itens do teste), dado que cada item pode contribuir com 1 (acerto) ou 0 (erro). Assim, o 0 expresso em termos de *x* corresponde à escala dos acertos, que é mais facilmente entendida pela grande maioria das pessoas.

Quanto às transformações não-lineares, deve-se dizer que há uma série delas; contudo, normalmente elas não facilitam a interpretação dos escores, o que, afinal, é a razão principal das transformações. De qualquer forma, a mais utilizada é a transformação *logits,* que faz uso de logaritmos naturais de base e. No caso do modelo de 1 parâmetro, ela seria

$$\ln \frac{P(\theta)}{Q(\theta)} = \theta - b \ ,$$

onde $Q(\theta) = 1 - P(\theta)$.

Entre as inúmeras possíveis aplicações da IRT na teoria dos testes, algumas são especialmente relevantes, nas quais a IRT tem contribuições inovadoras e promissoras.

Banco de itens

Na era do computador, a existência de banco de itens permite uma utilização não somente mais sofisticada, mas muito mais expediente, prática e eficiente da medida psicológica. A construção do banco de itens é viável dentro da teoria psicométrica clássica, fornecendo os parâmetros de dificuldade e discriminação de cada item. Entretanto, nesta teoria, os parâmetros são dependentes da amostra de sujeitos utilizada. A IRT permite estabelecer os mesmos parâmetros independentemente da amostra utilizada; daí é possível incluir sempre novos itens diretamente comparáveis com os já inclusos no banco. A técnica para esta façanha, entre outras, consiste em aplicar os novos itens juntamente com uma amostra de itens já incluídos no banco a uma amostra razoável de sujeitos e estimar os parâmetros dos novos itens em confronto com os dos itens utilizados do banco de itens. Assim os novos itens entram no banco nas mesmas condições que os velhos.

Testes a medida *(computerized adaptive testing* — CAT)

Anteriormente, foi visto que um teste fornece uma medida mais precisa da aptidão (0) do sujeito quando seus itens se situam no nível de dificuldade correspondente ao nível da aptidão do sujeito. Assim, um sujeito com G = 2 deveria ser examinado com itens de dificuldade *(b)* em torno de 2, dado que a função de informação deles, para tal 0, é máxima a esse nível. Na utilização tradicional dos testes, é costumeiro aplicar-se um mesmo teste a sujeitos de níveis diferentes de 0. Num caso destes, o teste avalia bem alguns sujeitos e mal outros. O ideal seria aplicar para cada sujeito um teste diferente, obviamente medindo a mesma aptidão, mas que se emparelhe, em termos de dificuldade, ao nível G de cada sujeito. Esta é a idéia atrás dos testes feitos a medida *(tailored* ou *computer adaptive testing).* Nesta situação, a seqüência dos itens submetidos ao examinando depende do desempenho do mesmo no item anterior; assim, para cada sujeito, a seqüência de itens é diferente de sujeito para sujeito. O que fica de problemático em tal procedimento é a garantia de que sujeitos diferentes submetidos a itens diferentes estejam sendo avaliados da mesma forma. A IRT, podendo estabelecer os parâmetros e as funções de informação dos itens, pode também demonstrar a equivalência entre testes diferentes, ou melhor, selecionar, para sujeitos diferentes, itens diferentes mas equivalentes.

Equiparação de escores *(test score equating)*

O problema que a equiparação dos escores procura resolver é o seguinte: se dois examinandos tomam testes diferentes para avaliar a mesma aptidão, é possível comparar os seus escores? Por exemplo, na seleção para entrada na universidade nos USA tomam-se decisões sobre quem é aprovado ou não baseadas em escores de testes diferentes para diferentes sujeitos. É isto justificável? Sim, se os escores são comparáveis. Este problema, que já vinha sendo tratado na psicometria clássica, particularmente desde o trabalho de Angoff (1971), recebeu solução bem mais condizente com a IRT. Se se dispõe de um banco de itens construídos através da IRT, então tem-se à disposição um elenco grande de itens cujos parâmetros são conhecidos. Para avaliar um nível qualquer de 0, é indiferente qual a amostra de itens utilizados, dado que o 0 independe do instrumento (elenco de itens) usado, conquanto se trate de itens que meçam a mesma aptidão, é claro, e cujos parâmetros sejam conhecidos. Quando se quer utilizar itens novos com sujeitos novos, é preciso que neste procedimento se incluam itens do banco, cujos parâmetros são conhecidos, para equiparar os parâmetros dos novos itens — façanha que é tipicamente viabilizada na IRT (veja **Banco de itens,** na página anterior).

Identificação de itens enviesados

A questão a que se refere este problema é a de justiça social. O teste não pode prejudicar um sujeito por causa de sua raça, sexo ou coisas similares (as minorias) que, em tese, não têm nada a ver com a aptidão sendo medida — problema levantado sobretudo por Jensen (1969, 1980). A questão em pauta se refere a que um mesmo instrumento produz medidas diferentes de uma mesma aptidão, porque entram, numa situação, fatores estranhos que não estão presentes na outra. Por exemplo: medir a temperatura da água fervente ao nível de mil metros de altitude sempre é diferente de quando medida ao nível do mar. Isto ocorre não porque a água ferve em graus diferentes segundo o termômetro, mas porque há, no caso, a presença da altura em relação ao nível do mar. Para comparar as duas medidas, é preciso controlar o fator altura. Várias técnicas foram propostas para tratar deste problema na psicometria clássica com referência aos testes (Scheuneman, 1979; Angoff, Ford, 1973;Angoff, 1982), mas foi a IRT que trouxe técnicas mais apuradas para esta tarefa. Basicamente, a técnica da IRT para corrigir a "injustiça" dos testes consiste em comparar as ICC resultantes do mesmo item ou teste para sujeitos diferentes e compará-las, para verificar qual o desvio das mesmas entre um grupo de sujeitos e outro. Assim, é possível estabelecer o funcionamento diferencial *{differential item functioning* - DIF) que cada item do teste tem com relação a estes grupos de sujeitos. O DIF é definido como: um item mostra DIF se sujeitos com a mesma aptidão 0, mas de grupos

diferentes, não têm a mesma probabilidade de acertar o item. O item que apresentar tal problema, o DIF, deverá ser revisto, refeito ou eliminado do teste. Há várias técnicas estatísticas para estabelecer o DIF (Hambleton, Swaminathan, Rogers, 1991; Ellis, 1991; Hambleton, Swaminathan, 1985).

CONCLUSÃO

AIRT não veio somente para ficar, mas constitui a teoria psicométrica predominante no dito Primeiro Mundo de hoje. Embora ela seja teoricamente complexa e praticamente exigente em seus procedimentos analíticos, parece imprescindível que todos os que trabalham com testes psicológicos tenham conhecimento da mesma e dela façam uso na elaboração de seus instrumentos. A literatura é abundante na área, e a disponibilidade de *softwares* apropriados o é igualmente. O desconhecimento da IRT no Brasil vem complicar ainda mais o lastimável estado atual dos instrumentos psicológicos aqui utilizados, dado que nem a psicometria clássica é ensinada adequadamente na grandíssima maioria das universidades brasileiras.

REFERÊNCIAS BIBLIOGRÁFICAS

ANGOFF, W.H. Scales, norms, and equivalent scores. In: THORNDIKE. R.L. (Ed.). *Educational measurement.* 2.ed. Washington, D.C.: American Council on Education, 1971. p.508-GC)0.

------------------. *Use of difficulty and discrimination indices for detecting test bias.* Baltimore (MD): The Johns Hopkins University Press. 1982.

ANGOFF, W.H., FORD, S.F. Item-race interaction on a test of scholastic aptitude. *Journal of Educational Measurement.* Washington, D.C.. v.10, p.95-105, 1973.

BAKER. F. B. Methodology review: item parameter estimation under the one-, two-, and three-parameter logistic models. *Applied Psychological Measurement.* St. Paul (MN). v.ll, p. 111-142, 1087.

BIRNBAUM, A. Some latent trait models and their use in inferring and examinee's ability. In: LORD, F.M.. NOVICK, M.R. *Statistical theories of mental test scores.* Reading (MA): Addison-Wesley, 1968. p. 17-20.

ELLIS, B.B. Item Response Theory: a tool for assessing the equivalence of translated tests. *Bulletin of the International Test Commission,* v.19 n.1/2, p.33-51, 1991.

HAMBLETON, R.K. Principles and selected applications of Item Response Theory. In: LINN, R.L. (Ed.). *Educational measurement* 3.ed. New York: McMillan. 1989. p.147-200.

HAMBLETON, R.K. (Ed.). *Applications of Item Response Theory.* Vancouver (BC): Educational Research Institute of British Columbia. 1983.

HAMBLETON. R.K.. SWAMINATHAN, H. *Item Response Theory,* principles and applications. Boston (MA): Kluwer, 1985.

HAMBLETON, R.K., SWAMINATHAN. H., ROGERS. H.J. *Fundamentals of Item Response Theory.* London: Sage, 1991.

JENSEN, A.R. *Bias in mental testing.* New York: Free Press, 1980.

------------------. How much can be boost IQ and scholastic achievement? *Harvard Educational Review,* Cambridge (MA), v.39, p.1-123, 1969.

LORD, F.M. An application of confidence intervals of maximum likelihood to the estimation of an examinee's ability. *Psychometrika.* v. 18, p.57-75.1953.

----------------- *.Applications of Item Response Theory to practical testing problems.* Hillsdale (NJ): Lawrence Erlbaum, 1980.

------------------. Standard error of measurement at different ability levels. *Journal of Educational Measurement,* Washington, D.C., v.21, p.239-243, 1984.

------------------. *A theory of test scores.* Iowa City (IA): Psychometric Society, 1952. (Psychometric Monograph, 7).

LORD, F.M., NOVICK, M.R. *Statistical theories of mental test scores.* Reading (MA): Addison-Wesley. 1968.

MISLEWY. R.J., BOCK, R.D. *BILOG:* maximum likelihood item analysis and test scoring logistic models. Mooresville (IN): Scientific Software, 1984.

MUÑIZ FERNANDEZ. J. *Teoría de respuesta a los ítems:* un nuevo enfoque en la evolución psicológica y educativa. Madrid: Piramide, 1990.

RASCH, G. *Probabilistic models for some intelligence and attainment tests.* Copenhagen: Danish Institute for Educational Research, 1960.

SCHEUNMAN, J. A method of assessing bias in the test items. *Journal of Educational Measure-ment.* Washington, D.C., v.16, n.3, p.143-152. 1979.

SWAMINATHAN, H. Parameter estimation in Item-response models. In: HAMBLETON, R.K. (Ed). *Applications of Item Response Theory.* Vancouver (BC): Educational Research Institute of British Columbia, 1983. p.24-44.

WINGERSKY. M.S., BARTON, M.A., LORD, F.M. *LOGIST users guide.* Princeton (NJ): Educational Testing Service, 1982.

WOODCOCK, R.W. *Development and standardization of the Woodcock-Johnson Psycho-Educational Battery.* Hingham (MA): Teaching Resources Corporation, 1978.

WRIGHT, B.D. Misunderstanding of the Rasch model. *Journal of Educational Measurement.* Washington, D.C., v. 14, p.219-226, 1977a.

------------------. Solving measurement problems with the Rasch model. *Journal of Educational Measurement.* Washington, D.C., v.14, p.97-116. 1977b.

WRIGHT. B.D., STONE, M.H. *Best test design.* Chicago (IL): MESA Press. 1979.

WRIGHT. B.D.. MEAD, R.J., BELL, S.R. *BICAL:* calibrating items with the Rasch model. Chicago (IL): University of Chicago. School of Education. 1979. (Statistical Laboratory Research Memorandum, 23B).

WRIGHT, B.D., PANCHAPAKESAN, N. A procedure for samplefree item analysis. *Educational and Psychological Measurement,* Durham (NC), v.29, p.23-48, 1969.

2ª PARTE

MÉTODOS E TÉCNICAS DE MEDIDA EM CIENCIAS DO COMPORTAMENTO

CAPÍTULO 8

CONSIDERAÇÕES EM TORNO DA MEDIDA DA INTELIGÊNCIA

Leandro S. Almeida
Universidade do Minho, Portugal

Se permanece atual na Psicologia e na sociedade em geral o estudo e a medida da inteligência, se muitas das nossas apreciações diárias e decisões decorrem de inferências que fazemos a propósito das nossas capacidades e das capacidades dos outros, certo que bastante dúvidas se vão acumulando no seio dos psicólogos quanto a um eventual consenso de perspectivas num futuro próximo a propósito da sua definição e medida. A polêmica em torno quer da definição quer da medida da inteligência tem permanecido ao longo do tempo, ultrapassando o período áureo da abordagem diferencial e dos testes (Almeida, 1988). Se muitas críticas foram formuladas a propósito dos autores fatoralistas da inteligência, apontando-se nomeadamente que descreviam produtos do desempenho e não a inteligência na sua essência e no seu exercício, certo que duas dezenas de anos de pesquisa no seio da Psicologia Cognitiva, já com estudos de incidência experimental, não têm conseguido maior uniformidade de posições no nível teórico e da prática. Se juntarmos a afirmação atribuída a Jensen (1969) — *intelligence, like electricity, is easier to measure than to define* — à possibilidade de que no virar do século, e apesar de tôdas as críticas, a avaliação da inteligência será ainda feita pelos testes tradicionais e nos mesmos moldes das provas com quase um século de existência (Horn, 1979), então podemos iniciar este texto dizendo quão difícil é falar sobre "medida da inteligência" ou pretender ter um discurso coerente e integrador das posições mais significativas no assunto.

Este capítulo reporta-se, sobretudo, ao conceito de inteligência como capacidade cognitiva no sentido do conceito de "inteligência geral". É oportuno lembrar também que as aptidões específicas não serão alvo deliberado deste texto e menos ainda a referência às medidas do rendimento *(achievement tests),* mais associadas aos conhecimentos e à aprendizagem.

Temos assistido nos últimos anos a uma diminuição do número de referências, de títulos de livros e artigos ou de comunicações em congressos usando as expressões tradicionais de "inteligência", "aptidões", "capacidade intelectual" ou "quociente de inteligência". Por outro lado, utilizam-se hoje mais conceitos substitutivos como "cognição", "processos cognitivos", "aptidão escolar", "aptidão cognitiva" ou "aptidão escolástica" (Reschly, 1990). Este fato, no entanto, não pode assumir-se como refletindo menor interesse pelo estudo da inteligência ou querendo traduzir alterações significativas quer na definição quer, e muito menos, na avaliação da inteligência. Por exemplo, a WISC-R, mantendo o termo "inteligência" na sua designação, interpreta o QI calculado como *school functioning level* (SFL); o *Lorge-Thorndike Intelligence Test* passou a designar-se *Cognitive Ability Test,* o *Otis-Lennon Mental Ability Test* designa-se agora *Otis-Lennon School Ability Test,* e na 4ª revisão da Stanford-Binet (1986) o tradicional QI passou a designar-se *Standard Age Scores* (SAS). As alterações, como se depreende, foram mais de "circunstância" que de "substância", muito embora poderão melhor enquadrar o uso dos testes e os seus resultados num movimento sócio-cultural que relativizou a sua objetividade, rigor e valor, ou em modelos da intervenção psicológica menos circunscritos à avaliação.

Como explicar tudo isto? Interessa ainda o estudo da inteligência? Serão ainda necessárias as medidas da inteligência?

Podemos enquadrar a resposta a esta questão fazendo referência a dois tipos de variáveis, aliás também importantes para explicar a diminuição do uso do termo "inteligência" ou a sua substituição por outros. A primeira tem a ver com o desencanto no seio dos psicólogos e da opinião pública com alguns excessos associados a meio século de uso e abuso da medida da inteligência (também da medida em geral na Psicologia). Poderiam os psicólogos não saber verdadeiramente o que os testes mediam, mas, desde que servissem para diferenciar e apresentar alguma correlação com o desempenho futuro dos sujeitos noutras situações, eram necessários e suficientes. Poder-se-ia não saber o que era a inteligência, mas, desde que se assumisse que o teste a avaliava, então ela passava a ser definida por "aquilo que o teste mede". Um verdadeiro ciclo vicioso se instalava na pesquisa e na prática. Empresas e recursos são maioritariamente conduzidos para a construção de novos testes, para a estandardização dos seus resultados, para o reforço da sua capacidade preditiva e para os estudos diferenciais, tomando as características sócio-culturais dos indivíduos.

Um segundo tipo de aspectos decorre da percepção hoje tida de que alguns construtos psicológicos necessitam ser melhor definidos tomando a sua essência e não apenas as suas manifestações externas.

Inclusive, algumas vêzes, questionamo-nos se algumas das variáveis psicológicas que estudamos — e isto não é exclusivo da inteligência — têm uma existência e substância próprias ou poderão não traduzir mais que meros artefatos e realidades criadas pelos psicólogos. Claro está que, mesmo sendo realidades criadas, elas podem ser igualmente úteis para descrever e analisar a realidade psicológica, ou seja, o comportamento humano. Contudo, nesse caso, temos que assumir uma postura diferente e mais humilde comparativamente àquela que nos caracterizou entre os anos 40 e 60 a propósito da medida da inteligência e das implicações práticas decorrentes dos seus resultados. Nessa altura, como os profissionais dos testes e, por meio destes, os analistas da mente, fomos afirmando não só a existência como a hereditariedade da inteligência, não só as diferenças interindividuais de capacidade como decidindo os projetos sócio-profissionais des sujeitos.

Dentro de uma postura mais relativista, podemos concordar com Anastasi (1986, p.5) que *the term intelligence has acquired too many excess meaning that obfuscate its nature.* Um pouco narcisicamente, os investigadores foram defendendo as suas posições, avolumando o número de dados a favor das suas teorias e justificando o interesse comercial das suas baterias de testes. Os psicólogos, na prática, pouco interessados no alcance teórico de tais discussões, acabavam por utilizar os testes que comercialmente se encontravam disponíveis. Com alguma intuição à mistura, eles acreditavam que tomando os resultados nos testes decidiam melhor da colocação profissional e da orientação vocacional dos indivíduos. Os testes ajudavam a cometer menos erros inferenciais na rotulação de uma criança como "deficiente mental" ou na fixação de um grau de demência a um adulto acusado judicialmente de um crime.

Basicamente, eram essas as condições do uso dos testes de inteligência. No contexto escolar importava definir em que medida a criança possuía as capacidades e os níveis de desenvolvimento intelectual requeridos para prosseguir uma escolaridade dita normal ou regular, ou qual o seu perfil de aptidões tendo em vista uma orientação escolar e profissional no quadro das alternativas sociais existentes. No contexto laborai, a avaliação da inteligência e das aptidões intelectuais foi amplamente usada em seleção profissional. Nas situações de clínica e de justiça, a avaliação da inteligência servia à definição do grau de consciência, de julgamento ou de crítica, a par de considerações sobre estados de demência ou de deterioração intelectual.

Permanecem atuais estes problemas? Permanecem atuais as formas de medida e o uso dado tradicionalmente à avaliação realizada?

Infelizmente, os problemas de insucesso nas aprendizagens escolares permanecem e algumas crianças apresentam dificuldades cognitivas acentuadas. Também — e aqui podemos dizer que felizmente —,

as alternativas de realização pessoal diversificam-se em termos escolares e profissionais, exigem-se níveis mais elevados de especialização de conhecimentos, ao mesmo tempo que nalgumas profissões se exige uma grande capacidade cíclica de reconversão profissional. Por tudo isso permanece necessário compreender as dificuldades de aprendizagem apresentadas por alguns alunos, orientar vocacionalmente os indivíduos e proceder a seleções no quadro do mundo do trabalho. Também ao nível da clínica e da justiça, os exemplos que atrás descrevemos continuam presentes na sociedade dos nossos dias e, possivelmente, justificando uma maior atenção por parte das autoridades e instituições. A evolução da sociedade não diminuiu a incidência de alguns problemas e dificuldades individuais; nalguns casos permitiu a sua mais livre manifestação, sensibilizou as pessoas para a sua existência ou, ainda, foi ocasião da sua maior freqüência.

Nao é possível diagnosticar tais problemas sem instrumentos adequados de medida, nem prevenir a sua manifestação sem a respectiva avaliação. Tampouco é possível realizarmos investigação nesses domínios sem bons instrumentos de medida. Se é verdade que para algumas situações de realização escolar e profissional algumas medidas mais diretas do desempenho parecem substituir as medidas mais diferidas, nomeadamente os testes tradicionais de inteligência e de aptidões, certo que estes últimos ocupam ainda um papel importante na avaliação psicológica. Por outro lado, se o desempenho da criança em tarefas escolares concretas ou a *performance* do adulto em determinadas atividades profissionais podem melhor servir o objetivo de uma avaliação contextualizada das capacidades de um e outro para aprender e para realizar, algumas vezes os testes psicológicos poderão ser necessários quando maior objetividade é reclamada na avaliação, quando interessa confrontar os desempenhos individuais com os resultados reportados a grupos de pertença (análise interindividual) ou quando importa recolher informação tendo em vista a estimação do desempenho futuro dos indivíduos. É possível, ainda, defender o uso dos testes como medida da inteligência menos contextualizada, ou mais na sua essência, ultrapassando-se mais facilmente experiências e conhecimentos singulares de cada indivíduo. Os objetivos da avaliação devem determinar a tipologia de instrumentos a usar, e é verdade que para algumas situações, onde antes se usavam os testes, recorre-se hoje a medidas mais diretas e mais contextuáis do desempenho cognitivo.

Por outro lado, podemos afirmar que a avaliação das aptidões cognitivas permanece hoje um dos domínios mais significativos da aplicação da Psicologia (Wigdor, Garner, 1982), mesmo que se tenha complementado os testes tradicionais com formas mais comportamentais de medida, como revelam os resultados de inquéritos junto aos psicólogos (Cruz, Almeida, Ribeiro, 1993; Murphy, Davidshofer, 1988), ou que se tenha

substituído a lógica seletiva anterior associada aos testes por uma outra tendo em vista o conhecimento e o aproveitamento de cada indivíduo (Tyler, 1981). Oakland e Hu (1992), tomando as respostas de psicólogos de 44 países, verificaram que, entre os dez testes mais usados internacionalmente junto a crianças e jovens, quatro deles dizem respeito a testes de inteligência geral (WÏSC/WISC-R na Iª posição; Raven na 2^8 posição; Stanford-Binet na 5ª posição; WAIS/WAIS-R na 6ª posição), dentre outros testes cognitivos (Bender, DAT, Frostig) e da personalidade (Rorschach, TAT, MMPI).

FORMAS DE AVALIAÇÃO DA INTELIGÊNCIA

Não faz sentido falarmos em formas de avaliação da inteligência sem primeiro abordarmos a sua definição. Aliás, ou fazemos esse esforço e conseguimos algum acordo entre os autores, ou todas as formas de medida que se proponham são no mínimo criticáveis. Julgamos que é importante interligar formas de avaliação a definições prévias, e minimamente aceitas, até para podermos interpretar convenientemente os resultados e saber diferenciar criticamente os diversos tipos de instrumentos disponíveis.

Duas abordagens da inteligência permaneceram na investigação e na prática psicológica em termos de definição e de medida da inteligência. Por um lado, a concepção de inteligência como uma entidade simples ou como entidade integradora de funções (vejam-se os conceitos de "inteligência geral", de "fator geral de inteligência", de "inteligência compòsita" ou de "quociente de inteligência"); por outro, a concepção de inteligência como uma estrutura de aptidões diferenciadas ou independentes entre si. Para o primeiro caso — aquele em que se situa este texto — podemos mencionar os trabalhos de Binet e de Wechsler, autores que não só definiram a inteligência como conjunto integrado de funções mentais — entidade unitária e global — como produziram baterias também elas reunindo itens diversificados tendo em vista a medida dessa globalidade funcional. Temos ainda autores como Spearman (1927), Vernon (1965) ou Cattell (1971) para quem a inteligência era sobretudo, ou em primeiro lugar em termos de importância, um fator geral entendido como capacidade de apreender e estabelecer relações, muito embora os dois últimos acrescentem fatores de segunda ordem ou de grupo. Se para Binet e Wechsler a inteligência seria melhor avaliada através de uma multiplicidade de tarefas cognitivas de índole verbal e prática, para Spearman ela sê-lo-ia através de tarefas abstratas, mais pautadas pela novidade que pelo conhecimento anterior do sujeito, e envolvendo o raciocínio indutivo e dedutivo.

Contrariamente a este primeiro grupo de autores, Thurstone (1938) e Guilford (1959) defenderam uma inteligência formada por aptidões independentes. No primeiro caso são propostas 7 aptidões diferenciadas, como compreensão verbal, fluencia verbal, aptidão numérica, velocidade perceptiva, aptidão espacial, memória e raciocínio (estas aptidões aparecem avaliadas através da bateria PMA do mesmo autor — *Primary Mental Abilities).* Guilford defende um número bastante mais elevado de aptidões, de acordo com o seu modelo *Structure-of-Intellect* (S.O.I.). Na versão mais recente do seu modelo, tomando a combinação de S operações mentais, de 5 conteúdos das tarefas e de 6 produtos possíveis do trabalho mental, chega à proposta de 150 aptidões diferentes (Guilford, 1982). Neste texto preocupa-nos a inteligência, definida no sentido de inteligência geral ou fator g, pelo que nos remeteremos exclusivamente ao primeiro grupo de autores.

A inteligência, dita inteligência geral, tem sido sobretudo avaliada através de tarefas envolvendo a compreensão, o raciocínio e a resolução de problemas, em conteúdos diversos (verbal, numérico, abstrato, espacial) e onde o que está mais particularmente em causa são os processos cognitivos associados ao relacionamento da informação e à aplicação de leis, princípios e relações inferidas. Por outras palavras, podemos afirmar que predominam aqui as funções cognitivas mais complexas comparativamente às funções simples de reconhecimento ou de memória, ou ainda às formas sensório-motoras predominantes nos testes de Galton e de MacKeen Cattell no começo do século (Almeida, 1988). Estudos numa orientação mais cognitiva de inteligência permitem-nos, de novo, reforçar a importância das componentes relacionais e inferenciais na sua descrição operacional (Sternberg, 1980).

A unanimidade dos autores em torno dos processos mentais mais valorizados e o recurso a materiais diversos na formulação das tarefas justificam que alguns autores falem numa inteligência verbal, numa inteligência prática ou numa inteligência abstrata. Jensen (1981) classifica os testes de inteligência geral em testes verbais (requerem o uso da linguagem, por exemplo provas de vocabulário ou analogias verbais), testes não-verbais (não requerem explicitamente o uso da linguagem, por exemplo os testes papel-lápis de matrizes, analogias ou seqüências figurativas) e os testes de realização (requerem o desenho do sujeito, a construção de algo ou a manipulação de materiais, por exemplo as provas de cubos ou de *puzzles).* As escalas de inteligência de Wechsler, por exemplo, recorrem a essa diversidade de situações.

A grande classificação que se pode assumir nos testes de inteligência geral, em nosso entender, é basicamente serem testes coletivos ou testes individuais. Veremos, então, várias especificidades destes testes tomando essa subdivisão.

Os testes de inteligência de aplicação coletiva datam da Iª Grande Guerra e do contributo dos psicólogos americanos na seleção dos recrutas. Já em 1923, no seu livro *Intelligence Testing,* Pintner identificava 37 testes coletivos de inteligência, entre os quais cinco não recorrendo à linguagem (tudo neles, inclusive as instruções, era figurativo) e seis não verbais.

São exemplos mais recentes de testes coletivos de inteligência: 1) o *Culture Fair Intelligence Tests* (Cattell, 1973), onde os itens, pretensamente avaliando a inteligência sem a interferência de fatores culturais, através do uso de figuras abstratas, cobrem diferentes tarefas — séries progressivas, classificação, matrizes e descoberta de propriedades comuns —, e 2) as *Raven's Progressive Matrices* (Raven, 1941, 1981, 1985), onde os itens são figuras geométricas na forma de matriz e onde falta uma parte a escolher entre 6-8 alternativas. Estas compreendem três versões devidamente diferenciadas: a) a *Coloured Progressive Matrices,* destinada a crianças; b) a versão estandardizada, composta por 60 itens; e c) a *Advanced Progressive Matrices,* formada por 36 itens e destinada a adultos mais desenvolvidos intelectualmente.

Algumas características gerais dos testes coletivos de inteligência podem apontar-se. Em primeiro lugar, são testes papel-lápis e, por norma, são realizados num tempo previamente definido. O limite de tempo não é, no entanto, fator decisivo no desempenho (trata-se de testes de competência e não de velocidade). Esse tempo limite serve sobretudo à busca de maior objetividade na avaliação e de maior diferenciação interindividual dos resultados. Nos testes de inteligência, como nos demais testes referenciados a normas, um procedimento usado na fixação do tempo limite num teste é tomar o tempo necessário para que 20% da população o possa concluir na globalidade dos seus itens.

Em segundo lugar, o formato dos itens nos testes coletivos de inteligência é geralmente de escolha múltipla *(multiple-choice),* cabendo ao sujeito assinalar a alternativa que lhe pareça mais correta. Sobretudo a partir dos testes coletivos usados na 1ª Grande Guerra, e com a generalização da avaliação da inteligência no campo da indústria e do ensino, a quase totalidade dos testes coletivos apresentam alternativas de resposta para escolha do sujeito. Este formato assegura uma maior objetividade na avaliação, nomeadamente no nível da escorização e cotação (permite, aliás, na maioria dos casos, uma informatização de tais procedimentos). O número de alternativas é variável, contudo procura-se que não seja inferior a quatro ou cinco, para reduzir a probabilidade de acerto meramente ocasional num item a 25% ou a 20%, respectivamente.

Em terceiro lugar, comparativamente aos testes individuais, é mais fácil a estandardização das condições de aplicação e de escorização dos resultados em testes coletivos. O aplicador pode estar menos presente na situação de avaliação, justificando isso inclusive uma certa tendência atual para a administração informatizada destes mesmos testes. Em termos de estandardização, podemos afirmar que, respeitando-se as instruções que constam dos manuais, e sendo a correção e escorização igual para todos os sujeitos, as diferenças encontradas nos resultados podem assumir-se, logicamente, como manifestação das capacidades individuais. Por último, a sua aplicação coletiva traduz, por norma, uma economia de recursos.

A sua principal dificuldade ou limitação tem a ver com a pouca atenção dada à forma como os sujeitos respondem ou a outros aspectos que poderiam servir a interpretação dos resultados, por exemplo as razões que levam um sujeito a escolher determinada alternativa não tida como correta pelo construtor do teste. Associada a esta limitação está uma outra que tem a ver com a percepção de uma discrepância entre o resultado final do sujeito (ou seja, o número de itens corretamente resolvidos) e a sua capacidade real em face da pouca interação entre o avaliador e o sujeito, ou seja, os poucos espaços para o confronto do sujeito com o seu desempenho. Por outro lado, o fornecimento das alternativas de resposta, ainda que favoreça a estandardização do *testing,* pode provocar o aparecimento de padrões específicos de resposta (raciocínio, comparação, justificação) ou também respostas ocasionais nem sempre devidamente ponderadas na análise dos desempenhos individuais. Apesar destas desvantagens, a informação disponível permite-nos afirmar que estes testes medem de fato, de alguma forma ou em certa medida, a inteligência (Jensen, 1980; Jackson, 1984).

ESPECIFICIDADES DOS TESTES INDIVIDUAIS

Os testes individuais de inteligência são hoje ainda bastante utilizados. Ou porque os testes coletivos podem sofrer de certa "crise" presente em face das reações suscitadas pelas avaliações massivas das populações escolares, ou porque os testes coletivos poderão não aprofundar a avaliação cognitiva requerida para certas decisões de classificação ou de seleção a tomar, certo que os testes individuais de inteligência permanecem bastante utilizados e a definir o principal papel dos psicólogos, por exemplo os psicólogos escolares nos Estados Unidos (Reschly, 1990).

São exemplos de testes individuais as bem conhecidas escalas de inteligência de Binet e de Wechsler. A Escala de Inteligência Binet-Simon (1905)foi objeto desucessivas reformulações nos Estados Unidos *(Escala de Inteligência Stanford-Binet;* Terman, 1916; Terman, Merrill, 1937, 1960)

e uma mais recente na França *(Nouvelle Echelle Métrique de l'Intelligence — NEMI;* Zazzo et al., 1966). Na sua revisão mais recente nos Estados Unidos, ou seja, a 4ª versão (Thorndike, Hagen, Sattler, 1986), a escala tenta integrar-se num modelo dito hierárquico de inteligência, onde, a par de um fator geral de inteligência, podemos falar em três fatores de grande grupo, como a aptidão cristalizada (raciocínio verbal e raciocínio quantitativo), a aptidão fluida ou analítica e a memória a curto prazo.

As escalas de inteligência de Wechsler, em termos de concepção de inteligência, não se diferenciam das de Binet (inteligência como conjunto de funções cognitivas). Aliás, contrastando com Spearman, refere que quanto mais se procura purificar os testes em termos de um fator geral mais este perde a sua capacidade avaliativa da própria inteligência. Esta não é apenas uma função mental simples, mas, antes, uma parte de um todo ainda maior, a personalidade (Wechsler, 1950). A principal diferença foi o abandono do Ql de razão e, logicamente, do conceito de idade mental, das vicissitudes do seu cálculo e a adoção do Ql de desvio.

A primeira escala construída designou-se *Wechsler-Bellevue Intelligence Scale* (Wechsler, 1939), que em 1955 passou a designar-se *Wechsler Adult Intelligence* Sca/e(WAIS) e da qual existe uma revisão mais recente (WAIS-R, 1981). Trata-se de uma escala destinada à avaliação da inteligência do adulto através de 11 subtestes — seis verbais (informação, memória de números, vocabulário, aritmética, compreensão e semelhanças) e cinco de realização (completamento de gravuras, cubos, *puzzles,* arranjo de imagens e código) — e que permite a obtenção, no final, de três quocientes de inteligência (Ql verbal, Ql de realização e Ql global). Duas outras escalas foram ainda desenvolvidas por Wechsler para a avaliação intelectual das crianças. Referimo-nos à *Wechsler Intelligence Scale for Children* (WISC), para as idades dos 6 aos 16 anos, e à *Wechsler Preschool and Primary Scale of Intelligence* (WPPSI), para as idades dos 4 aos 6 ou *6V2* anos. A WISC foi criada em 1949, tendo a sua primeira revisão em 1974(WISC-R). A WPPSI foi desenvolvida em 1968. Ambas as provas se assemelham, em termos de estrutura, aplicação, *scores* e interpretação, à WAIS.

Algumas características dos testes individuais de inteligência podem apontar-se. Nos testes individuais, para além da estandardização dos estímulos e das condições de *testing* que poderíamos afirmar serem também preocupação nos testes coletivos de inteligência, é importante considerar a relação interpessoal que se estabelece entre o avaliador (por norma um psicólogo) e o sujeito (criança ou adulto). Para além de nessa relação interferirem variáveis diversas do avaliador e do sujeito, como o sexo, a idade, a classe social ou a atração física, convém não esquecer que a própria situação de avaliação é já uma relação percepcionada de formas diversas pelos diferentes sujeitos e que tal percepção interfere no desempenho. Trabalhos clássicos confirmam, de fato, diferenças nos

desempenhos em função das percepções individuais da utilidade, da dificuldade e do sentido do *testing* (Roazzi, Spinillo, Almeida, 1991).

Como se tem afirmado, na avaliação individual, o sujeito percepciona-se a responder não aos itens ou ao teste, mas ao próprio psicólogo. Isso complica ainda mais a situação — por exemplo, as reações em face dos itens fáceis ou difíceis que a prova comporta. Por todas estas razões, é decisiva a relação criada na avaliação da inteligência através de testes individuais, devendo ser dada particular atenção aos primeiros momentos da avaliação. Mais uma vez, importa aqui assinalar que estes testes devem não apenas ser de aplicação exclusiva por psicólogos, como estes devem ter uma formação geral e específica nas técnicas usadas.

Comparativamente aos testes coletivos, nos testes de aplicação individual, por norma, não são fixados limites de tempo (o que não impede de se fixarem bonificações nos resultados para desempenhos mais rápidos). Por outro lado, é freqüente estes testes incluírem provas com manipulação de materiais e onde se procura uma informação avaliativa complementar ou alternativa à realizada através de itens verbais ou meramente figurativos. Finalmente, nos testes individuais, o sujeito não se limita a escolher uma alternativa, antes organiza uma resposta que será pontuada de acordo com o seu grau de adequação e de rigor em face da situação ou do problema apresentado. Aliás, na grande maioria destes testes, os sujeitos respondem a itens mais adequados ao seu nível de capacidade (evitam-se itens demasiado fáceis ou demasiado difíceis). Ao longo da aplicação das provas, o psicólogo, partindo das respostas corretas ou erradas anteriores, vai definindo que itens posteriores são passados ao sujeito.

Finalmente, os testes individuais são uma ocasião para a avaliação de aspectos cognitivos e não estritamente cognitivos do desempenho e importantes na interpretação dos resultados finais. O fato de a avaliação ocorrer numa situação individual, quase situação de "entrevista-clínica" ou "experimental-clínica", permite apreciar aspectos como: 1) a reação do sujeito ao *testinge* à sua prossecução (cooperação, ansiedade, bloqueios, persistência, etc), 2) o impacto de variáveis pessoais e situacionais no próprio desempenho (fadiga, motivação, autoconfiança, etc.) e 3) informação referente às respostas corretas e aos erros dados, designadamente em relação aos processos cognitivos aí usados. Consegue-se, assim, uma avaliação da inteligência mais aprofundada e interligando informações de índole quantitativa (pontos, notas, normas) e qualitativa (comportamentos, reações, relação, personalidade), utilizando-se os testes individuais em situações de diagnóstico e de seleção que requeiram uma maior exigência e individualização da informação. Possivelmente, em face da menor incidência da avaliação para efeitos de diagnóstico e de estimação das diferenças interindividuais, a favor de uma lógica mais preventiva e desenvolvimental, os testes individuais de inteligência

permitem uma avaliação mais compreensiva das capacidades dos sujeitos, nomeadamente tomando na análise do desempenho a interação entre capacidade e contexto ou tarefa (Roazzi, Spinillo, Almeida, 1991).

CONTROVÉRSIAS EM TORNO DA AVALIAÇÃO DA INTELIGÊNCIA

O aparente sucesso verificado no recrutamento por ocasião da ia. Grande Guerra através de testes de inteligência conduziu a um verdadeiro *boom* destes mesmos testes na sociedade civil a partir de então, nomeadamente nos Estados Unidos. Surge, então, uma grande indústria voltada para a avaliação em larga escala da inteligência (Murphy, Davidshofer, 1988), particularmente voltada para a satisfação de necessidades de maior objetividade e pretensa cientificidade das tarefas de seleção e classificação profissional.

No setor da Educação, e também nos Estados Unidos, esse *boom* teve lugar nos anos 60. O lançamento do Sputnik pela União Soviética despolutou um maior investimento do governo americano na educação, quer através de programas sócio-educativos voltados para a "compensação educativa" das crianças provenientes de meios sócio-culturais menos estimulantes, quer através da introdução nas escolas de uma prática sistemática de avaliação das capacidades intelectuais. De novo buscava-se a "excelência", e os testes voltaram a ser massivamente aplicados.

A controvérsia em torno dos testes, se era já uma realidade, ganhava logicamente particular incidência nesses momentos de utilização massiva. O uso desenfreado dos testes e a tomada dos seus resultados como absolutos suscitaram receios e reservas públicas. A partir de então esta controvérsia torna-se objeto de preocupações legislativas. Nalguns estados americanos, por exemplo, os testes de inteligência são pura e simplesmente abolidos, ou limitadas juridicamente as decisões com base nos resultados em testes.

Esta controvérsia assumiu diferentes argumentos, vários deles em torno do rigor e da validade das medidas de inteligência. Sem dúvida, os aspectos mais decisivos nesse debate público, inclusive no seio dos psicólogos, passavam pela questão da sua neutralidade social e pela questão do seu rigor científico. Em ambos os casos, o problema consistia na defesa/recusa de que um mesmo teste pudesse avaliar e comparar a inteligência dos indivíduos independentemente das suas origens e experiências sócio-culturais. Múltiplos argumentos foram usados contra os testes. Pondo de lado aqueles argumentos que mais poderão traduzir a animosidade do público, centrar-nos-emos nos que decorrem do baixo valor científico da medida da inteligência.

A neutralidade social dos testes de inteligência

Na segunda metade deste século, a contestação sócio-cultural aos testes de inteligência foi-se acentuando. Os grupos sociais mais desfavorecidos apresentavam resultados mais fracos nos testes tradicionais de inteligência e, por causa desse desempenho inferior, apareciam com maior freqüência segregados em termos de ensino e de emprego. Colocava-se, então, em causa a pretensa neutralidade social e validade científica dos testes. Falando em nome da Black Psychological Association, Jackson (1975) afirmava tal enviezamento nos testes e tomava-os como *quasi-scientific tool for perpetuating racism.* Claro está que esta mesma questão se pode generalizar a outras especificidades sócio-culturais dos individuos, que não apenas étnicas, nomeadamente quando específicas dos grupos sociais ditos minoritários.

Assim colocado o problema, podemos formar duas orientações na sua análise. Uma primeira que é apreciar os aspectos sociais que interferem na avaliação, dificultando desse modo a sua neutralidade social. Nesta altura, o esforço dos investigadores deveria ser dirigido para a concepção de formas de avaliação da inteligência menos afetáveis socialmente. Uma segunda orientação, mais radical, parte da afirmação da inteligência como uma construção social e, como tal, não pode ser avaliada fora do seu contexto. Neste último caso, a questão da neutralidade social da inteligência não faz qualquer sentido. Apenas respeitando as experiências sócio-culturais se poderá avaliar a inteligência, o mesmo é dizer que não podemos com um único teste ou tipo de testes avaliar a inteligência junto de qualquer grupo.

No quadro da primeira orientação, várias pesquisas procuraram diminuir a incidência dos fatores culturais na avaliação da inteligência. Umas centraram-se nos próprios testes e respectivos itens, enquanto outras procuraram cuidar das próprias condições de avaliação (situação de *testing).* Estão neste caso as tentativas de construção de testes livres de influências culturais (free *culture tests)* ou das experiências educacionais dos indivíduos (Eels et ai., 1951). Tais testes, de conteúdo figurativo-abstrato, permanecem, ainda hoje, como uma tentativa séria de diminuir — que não eliminar — a influência de tais variáveis na avaliação da inteligência e, ainda hoje, esse tipo de testes é usado em investigações transculturais como "um mal menor".

Para além do conteúdo e do formato dos itens nos testes de inteligência, em várias investigações questionou-se o impacto da própria relação avaliador-avaliado nos resultados de tais testes. A influência dessa relação no desempenho, e sendo essa influência diferente de acordo com as origens social e étnica dos sujeitos (Labov, 1970), veio reforçar o movimento *anti-testing.* As críticas foram-se tornando cada vez mais incisivas e abrangentes e, nalguns momentos, os testes foram associados

a práticas racistas (Jackson, 1975; Jones, Wilderson, 1976) ou aformas de desumanizar e de destruir crianças e jovens negros (Williams, 1970). Este último autor (Williams, 1972) construiu um teste em sua opinião melhor adequado à avaliação intelectual de crianças negras, pois que assente nas suas experiências culturais. Tentou demonstrar também que um teste de inteligência, enviesado culturalmente, poderia favorecer um grupo de crianças em relação a outros. Referimo-nos ao *Black Intelligence Test of Cultural Homogeneity* (BITCH), um teste de vocabulário assente nos conhecimentos culturais dos *ghettos* negros; nesta prova, as crianças negras obtêm melhores desempenhos que as brancas (Matarazzo, Wiens, 1977). Contudo, o fato dos itens estarem mais centrados em conteúdos específicos do que em processos cognitivos de tratamento da informação e de resolução de problemas questiona muito se de fato estamos perante medidas válidas de inteligência (Cronbach, 1978; Jensen, 1980; Matarazzo, Wiens, 1977), parecendo igualmente irrefutável que os sujeitos realizam melhor quando as tarefas dos testes se aproximam do seu quadro de referência cultural (Sternberg, 1991).

Permanece o problema e a controvérsia em seu redor. Será que os testes medem a mesma coisa em qualquer grupo humano? A questão pode ser mais básica ainda, ou seja: como pretender avaliar a inteligência — pelo menos parcialmente um produto cultural — pretendendo-se prescindir da sua origem? Será que alguma vez a avaliação da capacidade intelectual pode prescindir das experiências dos indivíduos? Não afetaremos os processos cognitivos avaliados ao pretendermos avaliá-los através de tarefas libertas dos conhecimentos culturais?

Os testes de inteligência, segundo alguns, seriam apenas fiéis e válidos para indivíduos da etnia branca e escolarizada (Williams, 1974). Com efeito, os sujeitos da etnia branca e escolarizada estão mais freqüentemente em contato com as atividades culturais, de resolução de problemas e de leitura e escrita que compõem as tarefas dos testes. Os testes cujos itens se aproximam ou refletem os conhecimentos e a experiência escolar não são socialmente neutros, pois tais aspectos não se encontram igualmente distribuídos. Os ¡tens verbais são aqueles mais facilmente afetados. Jensen (1980, p.637) acrescenta que os testes papel-lápis, requerendo a leitura, respostas escritas, velocidade, conteúdo verbal e evocação de informação antes aprendida são mais facilmente influenciados pelos aspectos culturais que, comparativamente, os testes de realização, figurativos, respostas orais, conteúdo não-verbal, potência e resolução de problemas. As implicações decorrentes destas considerações podem assumir alguma gravidade. O valor inquestionável dos testes pode perpetuar medidas socialmente discriminatórias. Os professores podem tender a explicar o baixo rendimento dalguns grupos de alunos tomando como referência o seu baixo desempenho em testes de QI *(cooling out* - Mercer, 1973). Ao mesmo tempo, a constatação de que

certos grupos são superiores e outros inferiores nos testes de inteligência pode sugerir uma determinação genética para tais diferenças (Jensen, 1969; Hernstein, 1971) e, em boa verdade, alguns setores da população pensam dessa forma. A questão da universalidade da medida e da comparabilidade interindividual dos resultados em testes de inteligência merece, no mínimo, maiores cuidados nas inferências e nas conclusões que os psicólogos produzem neste domínio.

A fidelidade e a validade dos resultados nos testes

Sem desmerecer o mérito e o rigor dos argumentos por detrás da controvérsia social antes apontada, mencionaremos agora alguns pontos que mais de perto parecem decorrer do estatuto "científico" da própria avaliação. Em primeiro lugar, é difícil assegurarmos que se avalia algo, ou convencer disso alguém, quando o próprio objeto da avaliação não se encontra ainda perfeitamente definido. Pelos menos em parte, esta é a situação ou a verdade a propósito da inteligência. Aliás, em algum momento avança-se num determinado sentido e mais tarde retoma-se o que havia sido abandonado. Como exemplo, volta-se hoje a reativar o interesse, já presente no começo do século, pela medida da inteligência tomando indicadores de natureza fisiológica (tempos de reação, potenciais evocados, condutibilidade elétrica da pele, etc). Por outras palavras, parece-nos que os testes serão sempre medidas possíveis da inteligência, mas nunca — pelo menos neste momento — "as medidas da inteligência".

A falta de uma definição clara e precisa de inteligência, capaz de reunir o consenso dos investigadores e estimular a produção de indicadores mais universais da inteligência, explica porque alguns investigadores e profissionais do terreno não se identificam com o tipo de testes existentes atualmente (cf. Sternberg, 1991); em sua opinião, os testes existentes valorizam determinados aspectos e não outros do desempenho cognitivo. Por exemplo, dão pouca atenção aos estilos cognitivos, à criatividade, às interfaces com o conhecimento dos indivíduos, às suas expectativas ou aos aspectos sócio-relacionais, favorecendo uma lógica "intelectiva" (Roazzi, Spinillo, Almeida, 1991).

Em nossa modesta opinião, a busca de estandardização e objetividade no material do teste (formato e conteúdo dos itens, procedimentos de resposta, tempo de execução, parâmetros de cotação, por exemplo) terão justificado um "artificialismo" nas situações apresentadas **e** na própria execução que, certamente, impede a todos os sujeitos a expressão das suas reais capacidades e, muito menos, a sua avaliação de uma forma mais compreensiva — por exemplo, como se organizam, funcionam ou manifestam nas situações concretas. Se alguns sujeitos se adaptam com facilidade às novas tarefas, ou mesmo ao "jogo" que se propõe, outros existem que apenas parcialmente se envolvem. Mais uma vez defendemos o teste como um instrumento ou meio de avaliação, e não

corno a avaliação no seu todo ou forma exclusiva de recolha da informação sobre a capacidade e o funcionamento cognitivo dos indivíduos.

É interessante assinalar que este "artificialismo" foi várias vezes justificado, quer para assegurar a objetividade da medida, quer para assegurar a sua própria neutralidade social. Procurou-se através dele diminuir o impacto das variáveis sócio-culturais e das experiências educativas na avaliação, seja de uma "inteligência pura", seja das diferenças interindividuais. Mais uma vez, aceitando-se a inteligência — pelo menos parcialmente — como um produto da socialização dos indivíduos, tal artificialismo deixa de ter sentido, pois, inclusive, dificulta a medida da inteligência na sua essência. Quando falamos no artificialismo dos itens nós estamos de algum modo a criticar a falta de ligação imediata entre o conteúdo, o formato e a linearidade dos processos subjacentes aos itens nos testes e conhecimentos anteriores ou as tarefas quotidianas dos indivíduos. Por norma, as tarefas nos testes são breves, simples e de execução linear, o que não corresponde à realidade quotidiana (Sternberg, 1985, p.30). Resta saber se, sendo isso necessário por razões de objetividade e de economia de recursos na avaliação, tais itens se constituem como elementos básicos (validade de conteúdo) e como bons indicadores (generalização) do desempenho habitual, ou se isso se aplica da mesma forma a todo e qualquer sujeito (universalidade). Aliás, o próprio "artificialismo" — veja-se a forte componente "abstrata" e "racional" dos itens — pode mais facilmente fazer parte das experiências de crianças e adultos da classe média, das sociedades mais industrializadas e escolarizadas.

Um outro aspecto prende-se com a melhoria dos desempenhos nos testes de inteligência com o treino ou a sua realização repetidas vezes. Para além de se questionar os ganhos daí decorrentes em face das definições de inteligência algumas vezes avançadas — por exemplo, a sua estabilidade e a sua independência em face da aprendizagem ou conhecimentos específicos —, podemos afirmar que tais ganhos afetam, no mínimo, ou a objetividade ou a validade dos resultados obtidos. A partir das instruções, os exemplos ou os exercícios de treino que são incluídos em cada teste deveriam, por norma, possibilitar a cada indivíduo a aprendizagem requerida para o desempenho de modo a este traduzir a sua "potência" (capacidade possuída). A melhoria verificada nos resultados com a repetição da realização de um teste, que não nos parece difícil de compreender ou explicar, questiona no imediato a idéia de que todos estejam à partida em igualdade de condições e, no final, sejam as diferenças observadas apenas interpretadas como refletindo diferenças reais nos níveis de capacidade possuída. Mais complicado, ainda, a relação que alguns estudos defendem entre as expectativas e as atribuições causais dos indivíduos (mais do foro sócio-motivacional) e os seus resultados nos

testes. Apesar da natureza essencialmente "intelectiva" destes testes, os resultados refletem também outras facetas dos indivíduos, ou seja, o impacto de variáveis mais ligadas à motivação e à personalidade (Barros, Almeida, 1991).

Outras dificuldades podem apontar-se, nomeadamente quanto à forma como são usados os testes. Algumas vezes eles não se encontram convenientemente adaptados e aferidos para os grupos populacionais onde aparecem usados, quer ao nível da investigação, quer da prática psicológica. Muitas vezes acaba-se por não respeitar nem atender às especificidades lingüísticas e culturais dos indivíduos (Reschly, 1981). Claro está que aqui o problema pode ser tido como anterior ao teste. Contudo também é verdade que alguns estudos sobre a validade preditiva dos testes partiram da formação prévia de grupos contrastantes de sujeitos. Por exemplo, em muitos trabalhos e na prática psicológica, a avaliação, por norma, ocorre já quando um problema de comportamento ou quando uma deficiência cognitiva se encontra identificada. Nessa altura não é difícil encontrar correlações mais ou menos elevadas entre os dois tipos de ocorrências. Em tais estudos, mais que estudar o poder prognóstico dos testes, reforça-se o sentido de diagnóstico do uso dos testes como se a ciência psicológica se resumisse, ou socialmente se justificasse, por essas atividades.

Finalmente, as condições do *testing* são diversas e decisivas para os resultados, e uma avaliação minimamente consciente requer uma individualização da informação. Esta individualização mostra-se particularmente importante em avaliações através de testes coletivos, pois que pouca informação o psicólogo pode obter a propósito das justificações e dos processos usados pelo sujeito nas suas respostas. Importa uma avaliação que permita uma imagem compreensiva e funcional do sujeito, o que em boa verdade nem sempre acontece na avaliação tradicional. Indivíduos com ótimos resultados nos testes nem sempre se apresentam socialmente eficientes, ao mesmo tempo que outros diagnosticados como "limítrofes" (entre a normalidade e a deficiência) se manifestam gênios em algumas áreas científicas e culturais. Algumas tentativas existem para esta avaliação compreensiva, mas são ainda mais a exceção do que a regra.

Uma das propostas mais conhecidas para uma avaliação abrangente e individualizante da inteligência é o programa *System of Multicultural Pluralistic Assessment* (SOMPA), de Mercer e Lewis (1978). O SOMPA destina-se a crianças entre os 5 e os 11 anos e integra a informação decorrente da 1) aplicação da escala WISC-R com a informação obtida através de 2) entrevistas aos pais, da 3) entrevista individual à criança, da 4) sua história médica, da 5) sua avaliação sensório-motora e perceptiva e da 6) aplicação de um Inventário de Comportamentos Adaptativos para Crianças. A informação referente ao *background* sócio-econômico deve

entrar numa equação de regressão para transformar o *score* de QI (WISC-R) num *score standard* reportado como Potencial de Aprendizagem Estimado *(Estimated Learning Potential* — ELP). A vantagem do SOMPA (dito programa de avaliação compreensiva da inteligência), segundo o seu próprio autor (Mercer, 1973; 1978), é o de permitir uma interpretação diferente para os *scores* obtidos (não a predição do sucesso esperado no futuro, mas o proveito que a criança poderá fazer das situações de aprendizagem que atendam ao seu contexto sócio-cultural) e o de não associar ao "estigma da deficiência" as percepções de incapacidade, dadas as conseqüências sociais e psicológicas nefastas que daí decorrem.

ALTERAÇÕES PREVISÍVEIS NA AVALIAÇÃO DA INTELIGÊNCIA

Algumas alterações são previsíveis, e em parte vão já sendo realidade, no que toca à avaliação da inteligência. Outras podem ter-se, para já como desejáveis num futuro próximo, em face da pertinência das críticas formuladas aos testes tradicionais de inteligência.

Uma primeira alteração podemos aqui tomá-la como esforço de maior pormenorização. Este esforço pode situar-se em diferentes níveis — por exemplo, tentando-se decompor conceitos clássicos de inteligência (aptidão verbal, quociente de inteligência, raciocínio) nos seus elementos processuais mais básicos e, com isto, não só "purificar" o teste como também o significado da sua informação. Um outro enfoque da pormenorização passa pelo maior interesse na análise intra-individual do desempenho no teste, em alternativa a uma apreciação reportada exclusivamente a normas de grupo (Perlman, Kaufman, 1990), sendo previsível o aparecimento de um maior número de provas referenciadas a critério *(criterion-referenced tests).* Este esforço de maior pormenorização permite antever uma informação qualitativamente superior, pois integra e atende aos processos cognitivos utilizados pelos sujeitos na obtenção dos seus *scores* globais, ao mesmo tempo que permite contextualizar essa informação nas experiências educativas, nos procedimentos de resolução de problemas ou nas estratégias e estilos cognitivos do sujeito em causa.

Uma segunda alteração prende-se com o alvo da avaliação. Aqui algumas mudanças parecem ser desejáveis, por exemplo uma valorização progressiva das capacidades e dos aspectos cognitivos mais diretamente ligados às aprendizagens e ao conhecimento, em desproveito de uma avaliação fortemente orientada para aptidão mental ou intelectual pura (Horn, 1979; Resnick, 1979; Sattler, 1988). Neste mesmo sentido vão as preocupações com avaliações intelectuais mais compreensivas, por exemplo integrando na avaliação da inteligência componentes não estritamente cognitivas (autoconceito, auto-eficácia, expectativas de

sucesso ou de fracasso), aliás, também não exclusivamente centradas nas aprendizagens curriculares (experiências e conhecimentos quotidianos). Espera-se que daqui possa resultar uma melhor compreensão quanto às formas habituais de adaptação e de resolução de problemas no quotidiano pessoal (Perlman, Kaufman, 1990).

Em terceiro lugar, a avaliação da inteligência deverá passar a incluir aspectos mais dinâmicos do próprio exercício e funcionamento cognitivo. Referimo-nos, por exemplo, aos estilos cognitivos ou às estratégias de processamento da informação ou, por outras palavras, não apenas e não tanto ao "quanto se é inteligente", mas ao "como se funciona cognitivamente". O teste K-ABC atende já a esta preocupação ao incluir provas para avaliação do processamento seqüencial e simultâneo de informação. Julgamos que este ponto de novo nos remete para a possível discrepância entre capacidade ou capacidades possuídas e desempenho manifesto nos testes e para a possibilidade dessa discrepância estar parcialmente associada aos procedimentos resolutivos menos adequados usados pelos sujeitos nas tarefas (cf. o conceito de "zona do desenvolvimento proximal" introduzido por Vygotsky, 1978).

Por último, o futuro da avaliação da inteligência deverá passar pelo impacto dos desenvolvimentos teóricos que hoje se estão a operar no campo da cognição (Lindemann, Matarazzo, 1990; Sternberg, 1991), nomeadamente os estudos experimentais no quadro da teoria do processamento da informação, a simulação dos processos mentais através do computador (ou o mero uso do computador na avaliação) e os estudos em psicofisiologia (compreensão dos fatores neurológicos, bioquímicos e fisiológicos relacionados com a atividade inteligente). Em primeiro lugar, os avanços registrados no campo da informática e da tecnologia em geral estão a ter algum impacto nas formas habituais de avaliação da inteligência. Se num primeiro momento apenas a escorização era feita de uma forma mecânica e rápida através destes meios, hoje a própria aplicação, organização de itens e elaboração de relatórios-síntese aparecem realizados através do computador. Potenciais vantagens deste procedimento têm a ver com 1) um conhecimento mais apurado das respostas aos itens isolados (tempo de latência, por exemplo), o que não é possível em testes coletivos papel-lápis, 2) uma aplicação mais de acordo com as reais características e capacidades dos sujeitos (eliminação dos itens anteriores (Weiss, Davidson, 1981) e 3) um menor tempo, menor custo e maior objetividade que daí podemos retirar. Evidentemente que estas vantagens não são absolutas. Importa verificar e controlar o uso, só por si, do computador. A familiaridade e a relação estabelecida com a máquina e o teste podem influenciar o desempenho, embora pareça já haver alguma informação de que em relação aos testes coletivos não se verifica uma perda significativa na acuidade dos resultados (Murphy, Davidshofer, 1988). Importa que os ganhos daí decorrentes em termos de

recursos humanos e materiais envolvidos não se convertam em perda de informação do *testing* e, conseqüentemente, do próprio sentido e relevância da avaliação psicológica (Perlman, Kaufman, 1990), muito embora tenhamos também que reconhecer — e isto infelizmente — que nem toda a avaliação anterior da inteligência atendia e integrava a diversidade de informações relativas à situação de teste.

Finalmente, os trabalhos experimentais analisando as correlações entre a inteligência e as tarefas cognitivas básicas, já muito próximas do substrato neurológico, permanecem ainda envoltos em grandes dúvidas quanto ao seu contributo efetivo para a avaliação da inteligência (Sternberg, 1991). Muito embora nas duas últimas décadas seja substancial o volume de investigações nesta área e se possa pensar que ultrapassamos já a falta de rigor nas medidas de Galton (Ceci, 1990), certo que é ainda bastante cedo para se concordar com autores que tomam os tempos de reação e outros indicadores básicos do processamento da informação como uma das medidas mais válidas e culturalmente neutras da inteligência (Eysenck, 1982).

CONCLUSÃO

Uma posição relativistica e crítica em relação aos testes justifica-se no presente, até para defesa dos próprios testes e do seu valor na prática psicológica. Diríamos que, olhando pelo prisma das suas características métricas, os testes de inteligência, sem se poderem tomar como rigorosos a 100%, garantem confiança e ajudam a tomar decisões. A margem de erro nas estimativas e nas decisões é menor com o seu uso do que sem ele. Evidentemente que esta margem de erro aumenta quanto mais próximo de um indivíduo — caso individual — se situarem a avaliação e as decisões. Ela é menor quando nos reportamos a grupos e particularmente a grandes grupos.

Reportando-nos aos testes mais referenciados na bibliografia ou usados na prática para avaliar as características intelectuais dos indivíduos, podemos afirmar que eles apresentam resultados satisfatórios em termos da sua fidelidade e validade. De um modo geral, os coeficientes de fidelidade assentes na estabilidade dos resultados e na consistência interna dos itens atingem valores acima de .75 e de .85, respectivamente. Alguma estabilidade de desempenho parece verificar-se ao longo da idade dos indivíduos, pelo menos entre a infância e o início da idade adulta (Rees, Palmer, 1970; Harnqvist, 1968; Wechsler, 1958; Matarazzo, 1972). Em vários estudos longitudinais se verificou, desde idades bastante precoces, uma relativa estabilidade nas medidas tradicionais do QI, nomeadamente já a partir dos 18/24 meses de idade (Bayley, 1949; Bradway, Robinson, 1961).

No nível da validade, podemos afirmar que os coeficientes de validade empírica e os parâmetros apreciativos da validade de conceito ou de construto ajustam-se também aos objetivos da avaliação psicológica prosseguidos em cada teste. Os testes de inteligência são ainda hoje medidas preditivas do sucesso do indivíduo na escola, no trabalho ou na vida social. Não dispomos, pelo menos até o presente, de substitutos adequados para os testes de inteligência quando o objetivo da avaliação psicológica se situa no conhecimento do nível e capacidade de funcionamento mental ou na predição do seu desempenho posterior. Claro está que, retomando algumas das críticas feitas aos testes, podemos estar aqui a tomar informação redundante. Por um lado, porque são tomados certos indicadores e não outros do sucesso e do fracasso em termos escolares e profissionais. Por outro, é previsível que os resultados nos testes se correlacionem com os desempenhos escolares na medida em que as tarefas apresentadas nos testes são similares às aprendizagens e atividades curriculares (Detterman, 1979). Isto não impede o valor informativo do teste e dos resultados obtidos em termos de diagnóstico e de previsão, mesmo que aceitemos quão bom seria se também o teste permitisse uma informação prescritiva, ou seja, dar informações relevantes para orientar a educação, o ensino e a aprendizagem (Perlman, Kaufman, 1990).

Os comentários produzidos em relação à fidelidade e à validade dos resultados nos testes não retiram a pertinência das críticas feitas ao método dos testes, nem tampouco podem significar que tudo está bem como está ou que tudo, desde que relativizado, fica resolvido. Vários problemas subsistem na prática. A generalização de resultados de amostras para subgrupos ou destes para os indivíduos singulares oferece particulares dificuldades, mesmo do ponto de vista deontológico. Por tudo isto, importa melhorar a qualidade dos instrumentos, cuidar do seu uso e precisar a interpretação dos seus resultados. Julgamos que as críticas jogam aí um importante papel pelo impacto que podem causar nos construtores e nos aplicadores dos testes.

Apesar das críticas, os testes permanecem apreciados e utilizados; não sendo perfeitos, oferecem no entanto maiores garantias e validade às decisões que se não existissem ou comparativamente a outras fontes alternativas de informação. Por exemplo, quando em alguns estados americanos os testes de inteligência foram substituídos por testes mais ligados ao rendimento da criança, uma maior percentagem de crianças provenientes das minorias étnicas continuaram a ser encaminhadas para os centros de educação especial (Vane, Motta, 1990). Informação acumulada em relação ao QI ou a outras medidas do desempenho cognitivo atestam que quem melhor realiza tais provas também obtém melhor rendimento noutros contextos de vida (Carroll, Horn, 1981).

Concomitantemente às críticas e à defesa dos testes, podemos também colocar o interesse com o estudo das diferenças individuais. Por

norma, um e outro aspecto andam juntos quando se referem às vantagens e às desvantagens da sua consideração na Psicologia. Mais uma vez parece-nos ser interessante para a prática psicológica o estudo das diferenças individuais. Para além dos aspectos comparativos daí decorrentes, e que nalgumas situações da prática psicológica são decisivos, mesmo nas relações de ajuda e de educação, o conhecimento de tais diferenças podem significar maior individualização. Atenda-se, por exemplo, ao paradigma *aptitude and treatment interactions* (ATI) e às suas aplicações no contexto escolar (Cronbach, Snow, 1977). Claro está que pretendemos aqui interligar "diferenças individuais" ao "esforço de individualização", o que nem sempre terá ocorrido. A este propósito podemos referir alguns desenvolvimentos na avaliação da inteligência decorrentes da aplicação dos conceitos de "zona de desenvolvimento proximal" (Vygotsky, 1978) e de "potencial de aprendizagem" (Feuerstein, 1979). O que está aqui em causa é sobretudo uma perspectiva dinâmica de avaliação, onde o avaliador interage, questiona, dirige e instrui o desempenho da criança, acreditando-se que, desta forma, se fica a conhecer melhor o seu funcionamento cognitivo e a discrepância entre a sua capacidade desenvolvida e a sua capacidade latente.

Este "esforço de individualização" é também decisivo numa avaliação psicológica, quando bem feita. A questão da falta de atenção ao indivíduo e ao seu contexto no *testing* é decisiva na avaliação psicológica, cabendo ao profissional zelar pela sua superação. A questão da "validade ecológica" pode melhor ser atendida se o psicólogo considera situações-problema como fonte dos testes. Também particularizando no caso da avaliação da inteligência, esta não pode ser reduzida à aplicação e escorização de um teste. Aliás, por mais perfeito que seja o teste, ou mesmo que se melhore a sua fidelidade e validade, certo que isso não torna o seu uso mais automático. O seu valor é sempre instrumental, e — como o nome indica — tudo depende da competência de quem o usa. Esse alguém deve ser melhor que os testes que usa (Kaufman, 1982, p. 13).

Na avaliação psicológica, a precisão do instrumento não dispensa a finura da intuição (Zazzo et al., 1978). O tomar apenas os testes ou seus resultados, por maior objetividade que detenham, não só é insuficiente como pode induzir-nos a erros. A par dos testes, importa considerar informação relativa a conhecimentos, atitudes e aptidões, por outras palavras, uma avaliação geral da competência (Sundberg, Snowden, Reynolds, 1978). Como refere Zazzo e colaboradores, "é preciso abandonar a idéia de que um processo se pode tornar tão preciso que possa ser confiado ao primeiro que surja; todo o processo científico não é mais que um instrumento que exige utilização por uma mão inteligente" (Zazzo et al., 1978, p.156). Isto tem sido, aliás, designado por *intelligent testing* (Kaufman, 1982; Wesman, 1968), se quisermos o uso inteligente dos testes de inteligência. As deficiências dos testes não colocam em causa a sua

utilização, contudo obrigam-nos a considerá-las quer na interpretação dos valores recolhidos quer na busca de informação complementar através de outros meios. Talvez por isso se justifique, e cada vez mais, a formação dos psicólogos na área da avaliação.

REFERÊNCIAS BIBLIOGRÁFICAS

ALMEIDA. L.S. *Teorias da inteligência.* Porto: Ed. Jornal de Psicologia, 1088.

ANASTASI, A. Evolving concepts of test validation. *Annual Review of Psychology,* v.37, p. 1-15, 1986.

---------------- . *Psychological testing.* New York: Macmillan, 1988.

BARROS. A.M.. ALMEIDA, L.S. Dimensões sócio-cognitivas do desempenho escolar. In: ALMEIDA. L.S. (Ed.). *Cognição e aprendizagem escolar.* Porto: Associação dos Psicólogos Portugueses, 1991.

BAYLEY, N. Consistency and variability in the growth of intelligence from birth to eighteen years. *Journal of Genetic Psychology,* Provincetown (MA), v.75, p.165-196, 1949.

BINET. A.. SIMON, T. Méthodes nouvelles pour le dignostic du niveau intellectuel des anormaux. *L'Année Psychologique,* Paris, v.ll, p.191-244, 1905.

BRADWAY. K.P., ROBINSON. N.M. Significant IQ changes in twenty-five years: afollow-up. *Journal of Educational Psychology.* Washington. DC, v.49, p.278-281, 1961.

CARROLL. J.B., HORN, J.L. On the scientific basis of ability testing. *American Psychologist,* Washington, D.C., v.36, p.1012-1020, 1981.

CATTELL. R.B. *Abilities, their structure, growth, and action.* Boston (MA): Houghton Mifflin, 1971.

---------------- . *Cattell culture fair intelligence tests.* Champaign (IL): Institute for Personality and Ability Testing, 1973.

CECI, S.J. *On intelligence... more or less:* a bioecological treatise on intellectual development. Englewood Cliffs (NJ): Prentice-Hall, 1990.

CRONBACH. L.J. Review of the BITCH Test *Eighth Mental Measurements Yearbook.* Highland Park (NJ): Gryphon, 1978. v.1.

CRONBACH, L.J.. SNOW. R.E. *Aptitudes and instructional methods-,* an handbook for research on interactions. New York: Irvington, 1977.

CRUZ. OR, ALMEIDA. L.S.. RIBEIRO, I.S. Avaliação psicológica e utilização dos testes em Portugal: comportamentos e atitudes dos psicólogos. In: ALMEIDA, L.S., RIBEIRO, I.S. (Eds.). *Avaliação psicológica:* formas e contextos. Braga: Associação dos Psicólogos Portugueses, 1993.

DETTERMAN, D.K. A job half done: the road to intelligence testing in the year 2000. In: STERNBERG. R.J., DETTERMAN, D.K. (Eds.). *Human intelligence.* Norwood (NJ): Ablex, 1979.

EELS. K. et al. *Intelligence and cultural differences.* Chicago (IL): Univ. Chicago Press, 1951.

EYSENCK. H.J. *A model for intelligence.* New York: Springer, 1982.

FEUERSTEIN, R. *The dynamic assessment of retarded performers,* the learning potential assessment device, theory, instruments, and techniques. Baltimore (MD): Univ. Park Press. 1979.

GUILFORD, J.P. Cognitive psychology's ambiguities: some suggested remedies. *Psychological Review,* Washington, D.C., v.89, p.4ª-59, 1982.

------------------. Three faces of intellect. *American Psychologist.* Washington, D.C., v. 14. p.469-479. 1959.

HARNQVIST, K. Relative changes in intelligence from 13 to 18. *Scandinavian Journal of Psychology,* Stockholm, v.9, p.50-82, 1968.

HERNSTEIN, R. IQ. *Atlantic-Monthly,* Boston (MA), p.43-64, Sept. 1971.

HORN. J.L. Trends in the measurement of intelligence. In: STERNBERG, R. J., DETTERMAN, D.K. (Eds.). *Human intelligence.* Norwood (NJ): Ablex, 1979.

JACKSON, D.A. *Multidimensional aptitude battery (MAB):* manual. Port Huron (MI): Research Psychologists Press, 1984.

JACKSON, G. On the report of the ad hoc committee on educational uses of tests with disadvantaged students. *American Psychologist.* Washington, D.C.. v.30. p.88-92. 1975.

JENSEN, A.R. *Bias in mental tests.* New York: Free Press, 1980.

------------------. How much can we boost IQ and scholastic achievement? *Harvard Educational Review,* Cambridge (MA), v.39, p.1-123. 1969.

------------------. *Straight talk about mental tests.* London: Methuen, 1981.

JONES, R., WILDERSON, F. Mainstreaming and the minority child: an overview of issues and a perspective. In: JONES, R. (Ed.). *Mainstreaming and the minority child.* Reston (VA): Council for Exceptional Children, 1976.

KAUFMAN, A.S. *Psicometria razonada con el WISC-R.* México: Manual Moderno, 1982.

LABOV, W. The logic of nonstandard English. In: WILLIAMS, F. (Ed.). *Language and poverty.* Chicago (IL): Markham. 1970.

LINDEMANN, J.E., MATARAZZO, J.D. Assessment of adult intelligence. In: GOLDSTEIN, G, HERSEN. M. (Eds.). *Handbook of psychological assessment.* New York: Pergamon, 1990.

MATARAZZO, J.D. *Wechsler's measurement and appraisal of adult intelligence.* Baltimore (MD): Williams and Wilkins, 1972.

MATARAZZO, J.D. WIENS. A.N. Black intelligence test of cultural homogeneity and Wechsler adult intelligence scale scores of black and white police applicants. *Journal of Applied Psychology,* Washington, D.C., v.62, p.57-63, 1977.

MERCER, J.R. *Labeling the mentally retarded.* Berkeley (CA): Univ. California Press, 1973.

------------------. Theoretical constructs of adaptive behavior: movement from a medical to a social-ecological perspective. In: COULTER, W.A., MORROW, H.W. (Eds.). *Adaptive behavior,* concepts and measurements. Orlando (FL): Grune and Stratton, 1978.

MERCER, J.R., LEWIS, J.R. *System of multicultural pluralistic assessment (SOMPA).* San Antonio (TX): Psychological Corporation, 1978.

MURPHY, K.R., DAVIDSHOFER, CO. *Psychological testing-,* principles and applications. [S.l.]: Prentice-Hall. 1988.

OAKLAND, T.. HU, S. The top 10 tests used with children and youth worldwide. *Bulletin of the International Test Commission,* v.19, n.l, p.99-120, 1992.

PERLMAN. M.D.. KAUFMAN. A.S. Assessment of child intelligence. In: GOLDSTEIN, G„ HERSEN, M. (Eds.), *Handbook of psychological assessment.* New York: Pergamon, 1990.

PINTNER, R. *Intelligence testing.* New York: Henry Holt, 1923.

RAVEN, J.C. *Manual for Raven's progressive matrices and Mill Hill vocabulary scales.* London: H. K. Lewis, 1981.

------------------ . Standardization of progressive matrices. *British Journal of Medical Psychology,* Leicester, v.19, p.137-150, 1941.

RAVEN, J.C., COURT, J.H., RAVEN, J. *Manual for Raven's progressive matrices and scales:* standard progressive matrices. London: H. K. Lewis, 1985.

REES, A.H., PALMER, F.H. Factors related to change in mental test performance. *Developmental Psychology Monographs,* v.3, n.2, 1970.

RESCHLY, D.J. Aptitude tests in educational classification and placement. In: GOLDSTEIN, G., HERSEN, M. (Eds.). *Handbook of psychological assessment.* New York: Pergamon, 1990.

------------------ . Evaluation of the effects of SOMPA Measures on classification of students as mildly retarded. *American Journal on Mental Deficiency,* v.86, p.16-20, 1981.

RESNICK, LB. The future of IQ testing in education. In: STERNBERG. R.J., DETTERMAN. D.K. (Eds.). *Human intelligence.* Norwood (NJ): Ablex, 1979.

ROAZZI, A., SPINILLO. A.G., ALMEIDA, L.S. Definição e avaliação da inteligência: limites e perspectivas. In: ALMEIDA, L.S. (Ed.). *Cognição e aprendizagem escolar.* Porto: Associação dos Psicólogos Portugueses, 1991.

ROSENTHAL. R.. JACOBSON. L. *Pygmalion in the classroom.* New York: Holt, Rinehart and Wiston, 1968.

SATTLER, J.M. Age effects on Wechsler adult intelligence scale-revised tests. *Journal of Consulting and Clinical Psychology.* Washington. D.C.. v.SO, p.785-786. 1982.

------------------ . *Assessment of children.* San Diego (CA): Jerome M. Sattler. 1988.

SPEARMAN, C. *The abilities of man.* New York: Macmillan, 1927.

STERNBERG, R.J. *Beyond IO:* a triarchic theory of human intelligence. Cambridge: Cambridge University Press, 1985.

------------------ . Death, taxes, and bad intelligence tests. *Intelligence,* Norwood (NJ), v. 15, p.257-269. 1991.

SUNDBERG, N.D., SNOWDEN, L.R., REYNOLDS, W.M. Toward assessment of personal competence and incompetence in life situations. *Annual Review of Psychology,* v.29, p.179-221.1978.

TERMAN, L.M. *The measurement of intelligence.* Boston (MA): Houghton Mifflin, 1916.

TERMAN, L.M., MERRILL, M.A. *Stanford-Binet intelligence* sca/e.3.ed.rev. Boston (MA): Houghton Mifflin, 1960.

------------------ . *Stanford-Binet intelligence scale.* Boston (MA): Houghton Mifflin, 1973.

THORNDIKE, R.L, HAGEN, E.P., SATTLER, J.M. *Stanford-Binet intelligence scale.* 4.ed. Chigago (IL): Riverside, 1986.

THURSTONE, L.L *Primary mental abilities.* Chicago (IL): Univ. Chicago Press, 1938.

TYLER, L.E. *Individuality,* human possibilities and personal choice in the psychological development of men and women. San Francisco (CA): Jossey-Bass, 1981.

VANE. J.R.. MOTTA. R.W. Group intelligence tests. In: GOLDSTEIN, G., HERSEN, M. (Eds.). *Handbook of psychological assessment.* New York: Pergamon, 1990.

VERNON, P.E. *The structure of human abilities.* London: Methuen, 1965.

VYGOTSKY. L.S. *Mind in society,* the development of higher psychological process. Cambridge (MA): Harvard University Press, 1978.

WECHSLER, D. Cognitive, conative, and non-intellective intelligence. *American Psychologist,* Washington, D.C., v. 15, p.78-83, 1950.

------------------ . *The measurement and appraisal of adult intelligence.* Baltimore (MD): Williams and Wilkins, 1958.

------------------ . *The measurement of adult intelligence scale.* Baltimore (MD): Williams and Wilkins, 1939 e 1944.

------------------ . *Wechsler adult intelligence scale-revised.* New York: Psychological Corporation, 1981.

. *Wechsler intelligence scale for children-revised.* New York: Psychological Corpora tion. 1974.

WEISS, D.J.. DAVIDSON, M.L Test theory and methods. *Annual Review ofPsychology.*.32, p.629-658, 1981.

WESMAN, AG. Intelligenttesting. *American Psychologist,* Washington, D.C., v.36. p. 112-113, 1968.

WIGDOR, A.K.. GARNER, W.R. *Ability testing:* uses, consequences, and controversies. Washington, D.C.: National Academy Press, 1982.

WILLIAMS, R.L *The BITCH test.* St. Louis (MO): Black Studies Program. Washington University, 1972.

------------------ . Danger: testing and dehumanizing the black child. *Clinical Child Psychology Newsletter,* v.9, p.5-6, 1970.

------------------ . The problem of match and mismatch in testing black children. In: MILLER, L. (Ed.). *The testing of black students:* a symposium. Englewood Cliffs (NJ): Prentice-Hall, 1974.

ZAZZO. R., GILLY M.. VERBA-RAD. M. *Nouvelle échelle métrique de l'intelligence.* Paris: Armand Colin, 1966.

------------------ . *Nova escala métrica da inteligencia.* Lisboa: Livros Horizonte, 1978. v.1.

CAPITULO 9

LA MEDICIÓN DE LA ORGANIZACIÓN LOGICA DEL PENSAMIENTO

Alfredo O. López Alonso
Universidad del Salvador
Buenos Aires. Argentina

Nuestra intención de medir la organización lògica del pensamiento puede remontarse al año 1979 cuando desarrollamos por primera vez el Test de Coherencia de Razonamiento, al que abreviadamente designaremos TCR. En aquellos momentos no éramos todavía conscientes — ni siquiera sospechábamos — de que con ese instrumento, creado expresamente para medir lo que su misma denominación indica — la coherencia — íbamos a poder evaluar más adelante un aspecto más amplio y complejo del pensamiento, como es su organización lógica.

Estábamos recién adentrándonos en nuestras investigaciones sobre procesos cognitivos e inferenciales, y hasta entonces el TCR era sólo un método de exploración. Teníamos una idea aún incompleta de lo que el test podía medir. Pero, si bien no estaba aún dentro de nuestras especulaciones la idea de medir la organización lógica del pensamiento, ésta fue una consecuencia directa y necesaria de los resultados que luego obtuvimos con ese instrumento.

Se dio de algún modo un ejemplo de lo que suele llamarse **serendipity,** pues un recurso metodológico como un test — generalmente producto de una teoría previamente elaborada — sirvió inesperadamente no sólo para ampliar esa teoría de base, sino también para generar una nueva teoría sobre un tópico prácticamente desconocido.

En síntesis, el TCR fue el factor de unión y generación entre una teoría previa sobre procesos de razonamiento y una teoría consecuente sobre la organización lógica del pensamiento.

Antes de pasar a detallar qué clase de instrumento es en sí el TCR, qué resultados brindó y qué otros aspectos del pensamiento permitió explorar — objetivo de este trabajo — vamos a referirnos al contexto y fines

El autor agradece profundamente a la Profesora María Herminia del Rey y a la Analista Programadora Viviana Nancy Agnoli, ambas de la Universidad del Salvador, por su valiosa e insustituible colaboración en los aspectos técnicos informático-computacionales.

teóricos dentro de los cuales se creó y desarrolló — contexto y fines que lo ligan ineludiblemente al estudio del pensamiento y la lógica.

RELACIONES ENTRE LOGICA. PENSAMIENTO Y PSICOLOGIA

La capacidad lógica de la mente humana ha sido siempre un tema de gran atracción y discusión por el gran poder de racionalidad que involucra y pone en juego.

Todo lo concerniente a esa capacidad despierta inmediatemente preguntas sobre qué es la lógica, qué es el pensamiento y cuáles son las relaciones entre una y otro. Y si a ello agregamos el supuesto de que existe una organización lógica natural del pensamiento, entonces se nos presenta el problema aun más complejo de cómo probar y definir esa organización y qué importancia reviste para cualquier actividad humana.

Las relaciones entre lógica y pensamiento son en cierto sentido paralelas a las de teoría y realidad; ambas están involucradas en toda argumentación humana tendiente a servir de fundamento a una estrategia probada de acción, toma de decisión o solución exitosa de problemas. De allí, que en un nivel profundo de análisis esas relaciones adquieran un carácter crucial.

En un libro clásico de enseñanza de la lógica, Francisco Romero y Eugenio Pucciarelli (1947) sostenían que ineludiblemente toda inteligencia dotada de una capacidad teórica tiene que sentirse atraída por la lógica, y veían a ésta como una manera en que el pensamiento y la inteligencia pueden comprenderse mejor a sí mismos. Aducían asimismo que los principios lógicos siempre han remitido a las leyes del pensar y del pensamiento — una asociación que se torna mucho más fuerte y explícita a partir de Boole (1847, 1854), un profesor de matemáticas inglés considerado el fundador de la lógica simbólica.

Boole desarrolló el álgebra de la lógica y del cálculo de clases como categorías esenciales de los procesos humanos de razonamiento y entendimiento. El primero de sus libros, *The Mathematical Analysis of Logic* (de 1847), estuvo dedicado a un análisis matemático de la lógica y al cálculo del razonamiento deductivo, y el segundo libro, *An Investigation of the Laws of Thought (de* 1854), estuvo destinado a investigar las leyes del pensamiento sobre las cuales él consideraba se fundaban las teorías matemáticas de la lógica y la probabilidad.

Con el programa de Boole se inicia un cambio revolucionario en la lógica y se aclaran mejor las relaciones entre ésta y el pensamiento. Para Boole la lógica es un sistema formal, un cálculo capaz de recibir distintas interpretaciones y cuya validez depende tan sólo de las leyes de combina-

ción de sus símbolos, cualquiera sea la interpretación que les demos. Su propósito era investigar las leyes del pensamiento y formalizarlas como tales. Inadvertidamente, éste era un propósito sumamente ambicioso que luego dio lugar a ciertas confusiones e interpretaciones contrapuestas sobre el tema, dado que algunos objetaron que Boole había confundido la lógica con las leyes del pensamiento, en tanto que otros entendieron que intentaba separarlas y aclarar su relación.

Uno de los investigadores actuales sobre el tema, Del Val (1977) quién suscribe en favor de la segunda interpretación, ha sostenido que "Boole fue, de hecho, capaz de separar la lógica de la psicología y de la epistemología, cosa que no habían hecho sus antecesores" (Del Val, 1977, p.20).

De nuestra parte, entendemos que no es apropiado confundir o identificar la lógica con las leyes del pensamiento, porque si bien aquélla surge como una formalización abstracta y explícita de las operaciones y procesos inferenciales de éste, nunca llega a abarcarlo en todos sus aspectos ni en toda su profundidad. Desde nuestro punto de vista, el pensamiento es siempre mucho más de lo que puede ser una lógica producto de aquél, ya que ésta sólo hará explícitos unos pocos de los múltiples y complejos procesos inferenciales implícitos del pensamiento y, por ello, dará solamente un "mapa somero" del mismo. De ahí que consideramos que el programa de Boole, aunque revolucionario e innovador para la lógica y su relación con el pensamiento, fue sumamente ambicioso por la amplitud y complejidad de los procesos que intentaba formalizar y sintetizar.

La lógica es una disciplina que vacía al pensamiento de sus contenidos y los estudia en cuanto a sus formas y a la generalidad de las mismas. Hace explícitas ciertas relaciones a fin de sustentar las razones necesarias y/o suficientes que deben esgrimirse para aceptar o rechazar concluyentemente la verdad o falsedad de unos argumentos — las consecuencias — a partir de la verdad o falsedad de otros — las premisas. La lógica, entonces, surge como una reflexión explícita sobre el proceso de pensar, a efectos de abstraer y generalizar algunas de sus formas y principios regulatorios, y justificar las conclusiones que se extraigan a través de los mismos, pero sólo a partir de su forma y valor de verdad, no de su contenido.

No obstante, la lógica no es una disciplina simple. Existe una lógica general que constituye la doctrina fundamental de toda lógica, y lógicas parciales y especializadas. Por ejemplo, junto a la lógica formal y asertiva de los valores de verdad existen también las lógicas modales — a ambas hemos de referirnos luego. Hay también una metalógica, desde que la lógica se constituye y se fundamenta recursivamente sobre niveles

progresivos de abstracción y formalización y el más elevado de esos niveles — el metalógico — provee los principios fundamentales y reglas formales de todo lenguaje lógico. De este modo, la lógica se constituye como ciencia, desligándose del pensamiento y de sus ataduras lingüísticas y psicológicas, como también de sus bases ontológicas, filosóficas o metafísicas, según se vea.

Si bien la lógica es un producto exclusivo del pensamiento, llega un punto en el que se independiza de éste, y se convierte en una instancia crítica, normativa, correctiva y evaluativa del mismo — dentro, por supuesto, de ciertos límites y condiciones previamente estipuladas. Vista de este modo, estimamos que la lógica cumple con un principio definido por Wundt (1886, 1889) como **principio de heterogénesls de los fines,** según el cual un proceso, producto de un sistema, se independiza de éste (en este caso sería la lógica respecto al pensamiento) y después pasa a constituirse en un factor externo de determinación y control del sistema original.

Aunque bajo otra denominación, este principio fue también adoptado por Vaihinger, filósofo positivista-idealista y biólogo pragmático, quién escribió hacia 1876-1877 una obra denominada *La Filosofía del Como Sí* en la cual explica cómo el conocimiento surge a base de una necesidad y utilidad biológica de adaptación del individuo al medio, y pasa luego a independizarse de la misma y a trascenderla.

De este modo, podemos esbozar como esquema a primera vista, que la lógica, aunque surgida del pensamiento, se separa necesariamente de éste como un producto explícito del mismo y pasa a regularlo externamente, como un conjunto independiente de principios normativos y de reglas de formación, transformación y control.

También se separa así la lógica de la psicología. Si bien la psicología se encarga de los procesos de pensamiento en cuanto tales, la lógica contribuye con la misma dando las formas, leyes y reglas de formación, y sintaxis, como condiciones generales para interpretar y evaluar esos procesos y sus productos.

La lógica contribuye así como una disciplina que puede establecer y evaluar la validez y correción formal y la eficiencia de esos procesos, pero sin explicarlos como tales. Por el contrario, su explicación como procesos le corresponde a la psicología.

Las leyes de la lógica constituyen entonces principios reguladores atemporales y de validez universal; atemporal¡dad y universalidad que prueban de algún modo su independencia del pensamiento y de la psicología, ya que éstos no gozan de esas mismas prerrogativas.

En síntesis, al pensamiento la lògica sólo lo justifica y valida formalmente en su corrección; en tanto que, la psicologia no lo valida pero lo explica como proceso, sea corredo o erròneo.

La reflexión sobre el pensamiento comenzó tempranamente entre los filósofos de la antigüedad. En primer lugar, Sócrates dio un método de autorreflexión, la "mayéutica", que nos permite darnos cuenta y hacer explícitos contenidos implícitos y no advertidos de nuestro pensamiento, que al hacerlos explícitos nos hacen reconocer nuestros errores y contradicciones.

Luego, Platón dio una doctrina de las ideas como entidades separadas de la realidad sensible y una primera interpretación del conocimiento a *priori* independiente de la experiencia. En tercer lugar, Aristóteles nos dio una doctrina explícita del razonamiento basada en su forma y en su valor de verdad, pero separada de su contenido o significado material; así como separó "materia" y "forma" en la constitución de nuestros conceptos. Con esto se pasó del pensamiento a la lógica.

Aunque Aristóteles utilizó el término "lógica" sólo como adjetivo, la primer sistematización de la lógica le pertenece, según trabajos reunidos bajo el nombre de *Organon.* El núcleo de su doctrina fue el "silogismo", término griego que en distintas acepciones significa "cálculo, cuenta, recapacitación, raciocinio, argumento demostrativo", en suma, "razonamiento" en sus aspectos más explícitos y autorreflexivos.

La lógica para Aristóteles fue básicamente instrumental, un método para separar el pensamiento válido del falaz y de este modo sustentar la veracidad de un nuevo conocimiento sobre la veracidad de nuestro conocimiento anterior. Para ello utilizó ciertas formas y expresiones del lenguaje — las proposiciones categóricas — en las cuales hizo explícitas la función de tres términos (sujeto-medio-predicado), ligados mediante la cópula (es - son) y términos que denotan la extensión del "sujeto", como "todos", "algunos", "ninguno" — si bien, en el siglo XIX, Bentham (1827) y Hamilton (1859) propusieron incluir expresiones de cuantificación para el término "predicado", ello no prosperó.

Las formas de los juicios categóricos son: "Todos los A son B" (universal afirmativa), "Ningún A es B" (universal negativa), "Algún A es B" (particular afirmativa), y "Algún A no es B" (particular negativa), las que se representan con los símbolos: A, E, I, O, respectivamente. De esta cuádruple categorización aristotélica, luego se derivó un cuadro de oposiciones, en el que se determinan ciertas compatibilidades e incompatibilidades entre esos cuatro juicios categóricos, según sean contrarios o subcontrarios, contradictorios o subalternos entre sí. A partir de ese cuadro de oposiciones se definen también distintas formas de inferencias inmediatas a las que apelaremos después.

Esto viene a colación de que la lógica, al tomar los aspectos formales, universales y atemporales, de los objetos del pensamiento, procura establecer demostrativamente el carácter necesario, implicatorio, compatible, contingente, excluyente o contradictorio, de las conclusiones que extrae de sus propios argumentos iniciales o premisas.

No obstante y como ya lo anticipáramos, hay lógicas especializadas que pueden tomar aspectos no tan puramente formales y asertivos de los objetos del pensamiento, como la "modalidad", la "necesidad", el "modo de ser", etc.. En tal sentido, por ejemplo, Georg Henrik Von Wright (1970) ha opuesto una "lógica modal" a la "lógica de la verdad" a la que nos hemos estado refiriendo preferentemente. Von Wright arguye que las lógicas modales hacen referencia a otros aspectos de las cosas distintas de la verdad, como su **necesariedad,** su **posibilidad** o **imposibilidad,** su **contingencia** (modos aléticos), su verificación, refutación o indecisión (modos epistémicos), su obligación, permiso o autorización, su prohibición o su indiferencia en tal sentido (modos deónticos), o bien su carácter universal (todos) o de existencia, inexistencia o vacuidad (algunos, ninguno) — modos existenciales.

La importancia de incluir, o tener al menos en cuenta, algunos de estos aspectos de la lógica modal viene al caso desde que — desde un punto de vista no sólo lógico, sino también psicológico — el modo en que está expresado un argumento, o bien los aspectos espontáneos y modales de una proposición, pueden alterar fundamentalmente la capacidad de los sujetos para comprenderlos y darles un significado unívoco y apropiado, y también para resolverlos de una manera lógicamente correcta. Por ejemplo, las proposiciones categóricas antes enunciadas de Aristóteles, incluyen términos como "todos", "algunos", "ninguno" que afectan a la extensión y existencia de los conceptos, y, por ende, su comprensión cabal. Creemos que a través de estos términos se vinculan aspectos de la lógica asertivo-formal de la verdad con aspectos de una lógica modal, lo que implica que los primeros no están totalmente exentos de los problemas de interpretación de los segundos. Este ha sido un aspecto clave tenido en cuenta en nuestro criterio exploratorio y evaluativo de la organización lógica del pensamiento a través del TCR.

Von Wright (1970) afirmaba que aunque los conceptos modales se distinguen de los conceptos de verdad, los dos dominios de categorías **no están del todo separados lógicamente,** ya que están ligados, por lo menos, por una relación asimétrica. Esta es: "si una proposición es **verdadera** entonces es posible ". Su conversa no es válida, ya que no todas las proposiciones posibles son verdaderas.

Si la **existencia y** la **extensión** de los términos de clase que integran las premisas de un razonamiento constituyen un factor modal que se suma a la afirmación y/o negación de los mismos para alterar las

transformaciones de la cópula (es - son), **entonces conviene ensayar distintas formulaciones a las ya dadas por la lógica clásica de predicados y silogismos.** Este es un aspecto al que volveremos más tarde cuando nos refiramos a las características formales del instrumento de evaluación lógica (el TCR) que expondremos en la próxima sección en sus aspectos lógicos y psicométricos.

La lógica de clases y su álgebra dadas por Boole y De Morgan fue continuada por Schroder (1890) y luego ampliada y sistematizada por Whitehead y Russell (1910) junto con la lógica de predicados, silogismos y funciones proposicionales. Paralelamente, la lógica fue analizada desde distintas perspectivas: por el empirismo (Hume, Stuart Mill), el criticismo (Kant), el formalismo (Peano, Hilbert), el enfoque gnoseológico-metodológico (Sigwart, Wundt), el logicismo (Zermelo, Dedekind, Frege, Russell y Whitehead), el intuicionismo (Brouwer) y el fenomenologismo lógico (Husserl, Pfãnder). También por "el empirismo lógico" (Wittgenstein, Carnap, Reichenbach) que puso el acento en el lenguaje y la verificación, y en la inducción y la probabilidad (Keynes).

Una variante más la constituyó el "psicologismo lógico", como el sustentado por ejemplo por Lipps (1893), quién sostenía que la lógica se apoya en lo psicológico y depende exclusivamente de esto. Esta posición ha sido criticada como una forma de reduccionismo psicologista y total relativismo, que desconoce la independencia de la lógica y de los principios básicos de la misma, como el de "identidad", "no-contradicçión" y "tercero excluido"; los que, si bien tienen una dimensión psicológica, son también principios ontológicos, ideales, metalógicos y gnoseológicos, separadamente de aquélla.

La lógica ha sido vista también como "ciencia normativa", sobre todo desde Kant (1800) y desde Herbart (1808,1813); es decir, como ciencia que prescribe formas y leyes del pensamiento. Este enfoque puede incurrir en el sesgo de sobrevalorar la lógica sobre el pensamiento, desconociendo el hecho fundamental de que la lógica, aunque norme y valide las formas y transformaciones del pensamiento, éste es siempre un **proceso viviente,** mucho más complejo que aquélla e indudablemente la desborda.

Si bien las relaciones entre lógica y pensamiento han sido abordadas por filósofos, lógicos y psicólogos, aún no están del todo claras y se estima que la coincidencia o base común que pueda haber entre los mismos, proviene, más que nada, de los principios ontológicos formales (de identidad, no-contradicción y tercero excluido) que fueron interpretados por el kantismo como principios trascendentes regulativos de la razón. Estos principios formales ontológicos son base de los principios lógicos y gnoseológicos homónimos, y sus distintos niveles no deben ser confundidos entre sí. Asimismo, Leibniz, uno de los principales filósofos del racionalismo, había ya sostenido que nuestro razonamiento se apoya en

dos grandes principios fundamentales: el de "no-contradicción" y el de "razón suficiente" como argumento de causalidad.

Por su parte, los psicólogos — como algo se adelantó — no han estudiado la lógica tanto como disciplina en sí, sino más bien como medio para ser contrastado con el pensamiento y estudiar de este modo el grado en que los sujetos razonantes — se entiende razonantes naturales no entrenados en lógica — son capaces de alcanzar espontáneamente o autónomamente los principios y estructuras formales de la lógica sin aún conocerlas. Si esta coincidencia se da, ello no sólo refuerza la universalidad de la lógica, sino también la racionalidad inherente del pensamiento humano y su natural organización lógica de base, aunque esto pueda suponer de algún modo un cierto *petitio principii,* a pesar de su independencia.

Pero aún así, a los psicólogos no sólo les ha importado establecer hasta qué punto el pensamiento humano es Isomórfico con la lógica, sino también, la medida en que no lo es, y se han preguntado, entonces, por qué desviaciones o fallas de la racionalidad ha de explicarse esa no-isomorfía; o bien, en última instancia, esa posible o eventual irracionalidad de la mente humana.

Para ello, tanto en un caso como en el otro, la lógica ha sido el referente y el juez ineludible y supremo. En tal sentido, la racionalidad o irracionalidad, como la logicidad y la ilogicidad del pensamiento humano, han sido y siguen siendo objeto de gran debate para la Psicología Cognitiva, lo que veremos en la sección siguiente.

EL PROBLEMA DE LA LOGICA DEL PENSAMIENTO DESDE EL PUNTO DE VISTA DE LA PSICOLOGIA COGNITIVA

El pensamiento puede verse como un complejo proceso de representaciones que pueden ser lógicamente consistentes o contradictorias; pero para adquirir status de conocimiento, esas representaciones, además de coherentes, deben ser un reflejo fiel y adecuado de la realidad. Esto involucra un problema epistémico en el cual no profundizaremos.

Las propiedades, posibilidades y fundamentación del conocimiento han sido tema clásico de la Gnoseologia o Teoría del Conocimiento, y parten de una relación básica entre sujeto cognoscente y objeto conocido. Esta cuestión es también hoy tema de las Ciencias Cognitivas, entre las cuales se encuentran la Psicología Cognitiva, la Psicolinguistica, la Inteligencia Artificial, la Neurociencia, etc. Pero, como ya dijéramos, aunque el conocimiento dependa de esa adecuación sujeto-objeto, la lógica prescinde de ella y sólo abstrae su forma, con lo cual el estudio del conocimiento no se completa. Por ello, la teoría del conocimiento y la

Psicología Cognitiva han debido aunar los aspectos puramente formales y los aspectos de contenido, como significado, conceptualización e inferencia, buscando una explicación más comprehensiva.

Dentro de las investigaciones psicognoseológicas del pensamiento ocupan un lugar destacado las investigaciones en psicología del razonamiento. Halpern (1990) define al razonamiento como el proceso basado en reglas que deciden lógicamente lo que tenemos que creer o aceptar concluyentemente. De este modo el razonamiento es el área de estudio de la Psicología Cognitiva que liga, a modo de puente, la lógica y el pensamiento.

En el estudio del razonamiento, se ha partido generalmente de la división clásica entre razonamiento inductivo y razonamiento deductivo. En el caso del razonamiento deductivo, las premisas o argumentos que tomamos como punto de partida suelen ser generalizaciones o, por lo menos, enunciados de una extensión mayor que la de la conclusión, abarcando o incluyendo a éstas en sus implicaciones. Estas generalizaciones suelen ser creencias sobre la naturaleza del mundo de las cuales extraemos instancias particulares como conclusión. El razonamiento deductivo tiene además el poder de contar con reglas de decisión (reglas de la lógica deductiva) por las cuales se puede demostrar la verdad de las conclusiones a partir de la verdad de las premisas, siempre que el razomaniento responda a una forma lógicamente válida.

Por el contrario, en el razonamiento inductivo las convicciones o creencias que tomamos como premisas o puntos de partida suelen ser hechos que hemos observado y verificado, de modo tal que a partir de ellos (un número finito y limitado de hechos) pretendemos extraer una conclusión, cuya extensión y nivel de generalidad es siempre mayor que la de las proposiciones que expresan aquellos hechos. Esta es la razón básica por la cual no podemos demostrar la verdad de las conclusiones a partir de la verdad de las premisas cuando usamos razonamientos inductivos. Es decir, éstos no son demostrativos.

En la práctica diaria, cuando las personas razonan no lo hacen de una manera que es pura o exclusivamente deductiva o inductiva, sino emplean alternativa e intercambiablemente ambos tipos de inferencia — además de otras. Sin embargo, resulta generalmente más verosímil suponer que es el razonamiento de tipo deductivo el que mejor refleja el grado de organización lógica del pensamiento, más de lo que permitiría hacerlo un razonamiento de tipo inductivo o probable, aunque éstos son también un reflejo de esa organización. Esta diferencia puede atribuirse al carácter concluyente, demostrativo, no-contingente y necesariamente no-contradictorio de las conclusiones de un razonamiento deductivo.

Uno de los problemas que más preocupó a los psicólogos cognitivos que estudiaban las relaciones entre lógica, conocimiento y pensamiento,

fue el de la consistencia o coherencia interna de sus procesos — esto significa su necesaria no-contradicción. En tal sentido, algunas contribuciones tempranas de autores de la Psicología Cognitiva social, como Feldman y otros (1966) y Abelson y otros (1968), dieron cuenta de que la consistencia del pensamiento constituía un elemento básico para explicar la organización del mismo. En la misma línea, Wyer (1974), al revisar distintos sistemas psicosociales de organización cognitiva, llegó a la conclusión de que todas las formulaciones teóricas sobre la manera en que dichos sistemas están estructurados, inferidos o interrelacionados, parten del común suspuesto de que los sujetos organizan sus conceptos, argumentos y creencias, de un modo internamente consistente o no-contradictorio.

También hay factores que ligan la consistencia interna con los sentimientos. En tal sentido, Festinger (1957) y Heider (1958) hallaron que la falta de consistencia en el orden cognitivo produce un estado displacentero en el orden emocional, el que atribuyen a una necesidad natural de los sujetos a eliminar sus inconsistencias o contradicciones internas. De este modo la consistencia puede verse también como una forma de equilibrio a ser logrado entre sus procesos intelectuales y emocionales.

En un contexto teórico vinculado, McGuire (1960), por un lado, y el mismo Wyer (1974) por el otro, han dado una explicación alternativa de la consistencia cognitiva y su logro. Ambos han hecho especial referencia a un efecto organizacional, denominado "efecto socrático", según el cual los sujetos tienden a incrementar espontáneamente su consistencia interna a medida que van tomando conciencia de sus contradicciones y al constrastar sus distintas convicciones y creencias en plazos relativamente breves de tiempo. Wyer destaca especialmente este efecto de *timing* para explicar procesos cognitivos de cambio y organización más amplios y complejos. Asimismo, McGuire, quien ha formulado una teoría psicosocial de la organización cognitiva, no sólo toma aspectos puramente formales o lógicos de la misma, sino también destaca similaridades y diferencias entre principios lógicos, tomados como paradigma o modelos de razonamiento, y la conducta real, vital, de procesamiento de la información que llevan a cabo los sujetos permanentemente. En tal sentido McGuire ha tratado de comparar el "razonamiento formal" con el "pensamiento deseado" *(wishful thinking)* en relación con procesos de cambio de actitudes y un estado de "armonía interna" que los seres humanos necesitan mantener a modo de equilibrio, entre sus sistemas de acción, creencia y pensamiento.

A su vez, Abelson y Rosenberg (1968) siguiendo la misma línea de investigaciones en consistencia cognitiva, y partiendo del aspecto dual, lógico y psicológico a la vez, de esta problemática, han acuñado el concepto de "psico-lógica" para explicar el equilibrio interno en el cambio y organización de actitudes y para describir cómo obran factores lógicos y extralógicos de equilibrio interno en la conducta social de los individuos. Consecuente con estos argumentos, el mismo MacGuire distinguió, en la

dinámica de la conducta y de las representaciones psicosociales, la intereacción existente entre una consistencia lógica y otra psicológica, admitiendo que ésta última, si bien es fuertemente demandante, no es tan estricta, precisa, ni estable como la primera. Finalmente, en la misma línea psicosocial, Wyer concluyó que las cogniciones y el procesamiento inferencial de la información tiende a organizarse de acuerdo con ciertas leyes que deben ser exploradas y expresadas formalmente a fin de explicar y justificar su consistencia interna.

De estos autores se extrae la conclusión de que en la dinámica actitudinal e Inferencial de los seres humanos, como asimismo en sus argumentos, decisiones y acción razonada, existe una pugna e interacción permanente entre argumentos lógico-formales y necesidades extralógicas y/o psicológicas. Aún más, ciertos autores como Lord y otros (1979), Kassin (1979) y Wason (1966-1968) han descripto ciertos efectos diferenciales, en el sentido de que los sujetos parecen ser más sensibles y consecuentes con la lógica, según ésta tienda a confirmar, más que a desconfirmar, sus actuales creencias y expectativas; en tanto, que tienden a racionalizar o distorsionar confirmatoriamente la información cuando resulta lógicamente refutatoria de las mismas.

Otros autores, como Lindzey y Aronson (1969) han sostenido que las personas, en última instancia, se comportan de una manera a través de la cual procuran elevar al máximo la consistencia dentro de su sistema lògico-cognitivo, y entre éste y su conducta actitudinal, pública y abierta.

Si bien la consistencia lógica-psicológica de la que hablan estos autores no es la misma consistencia interna lógico-formal del pensamiento que estamos investigando, ambas se ligan entre sí y hallan un común denominador a través de la idea de isomorfismo entre lógica y razonamiento natural o social.

Los argumentos esbozados por dichos investigadores psicosociales sugieren una idea importante para sustentar nuestra tesis fundamental del isomorfismo: la idea de que los aspectos racionales (lógicos) e irracionales (ilógicos) del pensamiento interactúan entre sí procurando alcanzar el nivel más alto de consistencia interna (o el número más bajo de autocontradicciones) antes de eludir la "indeseada refutación" o de aceptar la "deseada confirmación".

Estas cuestiones han generado una prolongada controversia entre dos corrientes principales y antagónicas de la teoría psicológica del razonamiento a las que cabe referirnos sumariamente: una, que sustenta aprioristicamente la logicidad y racionalidad esencial de la mente humana; y otra que supone una suerte de ilogicidad o irracionalidad sistemática. Esta controversia y su posible solución están íntimamente ligadas a los objetivos del TCR (López Alonso, 1990, 1991).

En torno a esa cuestión y en relación con el área específica del razonamiento silogístico, Mayer y Revlin (1978) han destacado tres líneas básicas de investigación: a) una línea destinada a indagar la naturaleza de los procesos inferenciales — línea en la cual nos insertamos; b) una línea que implementa los silogismos a modo de estándar diagnóstico contra el cual comparar la ejecución de diferentes grupos en relación a nivel intelectual — por ejemplo Guilford (1959), Thurstone (1938), Frandsen y Holder (1969) — o bien en relación a diagnóstico clínico-diferencial — por ejemplo, Feather (1965), Gottesman y Chapman (1960), Von Domarus (1944) — y, finalmente, c) una línea de estudios que adopta el silogismo en su sentido lógico clásico como modelo de pensamiento, línea dentro de la cual incluye a Boole (1854).

En la línea de investigación a la que estimamos pertenecer — la línea a) — el acento se ha puesto tanto en la codificación de los materiales inferenciales por parte del sujeto como en las reglas de inferencia que emplea o pueden utilizarse como contraste o comparación. En esta línea, por ejemplo, se insertan los trabajos de Rimoldi (1955, 1984) en los que compara la "táctica de resolución de problemas" empleada por un sujeto con una "táctica ideal" sustentada en la lógica y en el empleo lógico de la información. Pero dentro de esta línea siempre ha sido una preocupación fundamental la determinación y contrastación de los factores lógicos, por un lado, y psicológicos o extralógicos, por el otro, los que inducen a cambios significativos, a variabilidad y a un rendimiento diferencial en la capacidad y organización lógica de los sujetos.

De cuestiones emanadas de esta línea de investigaciones, surgió la discusión sobre la racionalidad o irracionalidad del pensamiento, según la forma en que los sujetos seleccionan, construyen, representan o comparan sus materiales inferenciales y resuelven sus problemas. Dentro de esta línea están quienes dan mayor importancia a la comprensión cabal de esos materiales por parte del sujeto, como también a su habilidad para almacenarla y recuperarla organizacionalmente en el momento oportuno. Esta es una línea que puede ser incluida dentro del enfoque que sustenta la racionalidad básica — lógica — del entendimiento humano. Dentro de esta línea se encuentran autores como Richter (1957), Chapman y Chapman (1959), Henle (1962,1978), Smedslund (1970), Ceraso y Provitera (1971), Erickson (1974, 1978), Revlin (1975), Dickstein (1976), etc.

Paralelamente, se encuentra en la misma línea el enfoque opuesto al anterior, aquél que supone la existencia de ciertas "formas" de inferencia que son erróneas o carentes de racionalidad, pero que se reiteran de manera sistemática en la conducta humana de resolución de problemas. Esta tendencia ha intentado explicar, por ejemplo, los errores o fallas en la resolución de problemas silogísticos y de probabilidad subjetiva, por cuenta de factores extralógicos o sesgos heurísticos sistemáticos y no-racionales.

Dentro de esta línea se enrolan autores como Woodworth y Sells (1935), Sells (1936), Morgan y Morton (1944), Frase (1966), Lefford (1946), Wasori (1984), Evans (1983), Pollard (1982), Johnson-Laird (1983, 1990), Tversky y Kahneman (1974) y otros.

A pesar del antagonismo de las dos líneas arriba mencionadas, ambas abocadas al estudio de los procesos inferenciales de problemas lógicos, creemos que el estudio de dichos procesos, desde la perspectiva del **Isomorfismo entre lógica y razonamiento** y desde el grado en que se da diferencialmente en el sujeto — tal como pretendemos lograrlo mediante el empleo del TCR —, permite dirimir esa cuestión básica entre racionalidad-logicidad e irracionalidad-ilogicidad. Confiamos en que la medición del isomorfismo mencionado, además de dar un índice válido y gradual de la organización lógica del pensamiento, constituye un recurso apropiado para superar y trascender dicha disyuntiva. Esta problemática y su relación con el TCR puede verse en López Alonso 1988, 1990, 1991.

Allí el TCR está dirigido a estudiar la capacidad natural diferencial del razonamiento deductivo del ser humano, contrastando las premisas y conclusiones de un razonador no-entrenado, con las conclusiones que pueden extraerse de esas mismas premisas cuando se aplican las reglas y principios de la lógica de conjuntos — y hallando, de este modo, el **grado de isomorfismo entre razonamiento y lógica.**

En los razonamientos deductivos, la manera en que los sujetos formulan o codifican sus premisas incide directamente en la corrección o no-contradicción de las conclusiones. Este argumento ha sido sostenido por Mary Henle (1962, 1978) y la gente que ha seguido su línea de trabajo, y ha sido experimentalmente probado por el autor y sus colaboradores usando el TCR (López Alonso, 1991, 1992; Ricardi, 1992; López Alonso y Ricardi, 1993). Por tal motivo, hemos adoptado el criterio de tipificar y computar los errores de contradicción en relación con la estructura de premisas como **criterio diferencial** para evaluar la organización lógica del pensamiento.

Con dicho instrumento, el razonamiento deductivo producto de una larga cadena de inferencias hechas por el sujeto, y que careciera totalmente de errores de contradicción, estaría reflejando un excelente nivel de organización lógica del pensamiento de dicho sujeto. Contrariamente, en la medida en que haya un número mayor y creciente de errores de ese tipo, ello estaría reflejando un grado progresivo de menor organización hasta llegar a un punto de desorganización o incoherencia total.

A propósito de lo que venimos exponiendo, Halpern (1990) ha sostenido que en el razonamiento cotidiano uno de los requisitos que la gente espera ver satisfecho para su aceptación (sea deductivo, inductivo o una mezcla de ambos) es que las premisas, además de verosímiles, les

resulten consistentes. Esto es importante porque son justamente las premisas las que van a sustentar las conclusiones de todo razonamiento y de toda toma de decisiones o soluciones de problemas, y lo que queda a *posteriori* es sólo la consistencia o coherencia que las conclusiones o decisiones deben guardar con las mismas. Estas consideraciones no sólo reflejan la importancia del rol de las premisas en el razonamiento y en la toma de decisiones, sino también la importancia de la consistencia interna de todo proceso inferencial en relación con aquéllas.

Pero otro punto de importancia que se desprende de lo anterior es evaluar hasta qué punto la gente es capaz de preveer o detectar **con suficiente anticipación** el riesgo de cometer contradicciones en el desarrollo de su propio pensamiento. Esta capacidad de prevención o anticipación no sólo es tomada como un signo de organización lógica del pensamiento, sino también como un requisito de importancia fundamental para la eficiencia y confiabilidad del pensamiento en general y del razonamiento en particular, tanto en sus aspectos prácticos de aplicación cotidiana que eviten el error u otros riesgos, como en sus aspectos formales, más abstractos y generales.

Con todas estas consideraciones y antecedentes teóricos referidos al contexto lógica/pensamiento-razonamiento/psicología en torno a la coherencia o consistencia interna, pasamos a ver en detalle los aspectos técnicos, lógicos y psicométricos del instrumento en cuestión — el TCR.

METODO

Descripción de la técnica (TCR), de su hoja de respuesta y primeras definiciones lógico-conjuntisticas

El instrumento con el cual se ha intentado medir directamente la coherencia interna de razonamiento o su no-contradicción, e indirectamente la organización lógica, implícita, del pensamiento, es el antes aludido Test de Coherencia de Razonamiento (TCR). El TCR fue diseñado por primera vez por el autor en 1979. En ese momento se definieron dos formas y subformas (Formas la, Ib, Ha, lib) que fueron descriptas por López Alonso (1981, 1988) y por Orsi (1988).

El TCR consiste de dos partes. Estas dos partes son claramente presentadas en la hoja de respuetas del TCR, un ejemplo de la cual es dada en la Figura 9-1. Como allí puede apreciarse, en la primera parte, se le da al sujeto una lista de cuatro términos o atributos relacionares entre sí, con una breve y concisa definición. Estos cuatro atributos son intercambiables y según cuáles sean los que se adopten, se define con ellos la forma del test. Cualquiera sea la forma, esos cuatro atributos se representan con las letras A, B, C y D. Por ejemplo, la forma I (subformas a,b) se

construyó con los atributos **Respetable (A), Honorable (B), Admirable (C), y Venerable** (D), como términos de valoración ética y social de las personas a ser relacionados entre sí. La forma II (subformas a,b) se construyó con los atributos **Americano** (continental) (A), **Argentino (B), Europeo (C),** e **Inglés** (D) dados como lugares de nacimiento de las personas — país o continente. A modo de ejemplo, en la Figura 9-1 se da un modelo de la hoja de respuesta de la Forma IIa del TCR.

Forma II a

Coherencia: Significa que dos afirmaciones son compatibles o no-contradictorias.
PRIMERA PARTE: Estime los siguientes porcentajes de acuerdo con su criterio (entre Oy 100)

- De toda la población mundial, qué porcentaje es AMERICANA .... 60. .... %
- De toda la población mundial, qué porcentaje es ARGENTINA .... 55.......... %
- De toda la población mundial, que porcentaje es EUROPEA .... 30. .... %
- De toda la población mundial, que porcentaje es INGLESA .... 25. .... %

Le agradeceremos que complete los siguientes datos:

NOMBRE: .... XX .... Código Nº .... 1028-D ....
LUGAR DE NACIMIENTO: .... B. Aires .... SEXO: .... Masculino ....
FECHA DE NACIMIENTO: .... 5 agosto 1969 .... EDAD: .... 22 ....
INSTITUCION: .... XX .... NIVEL ESCOLAR: .... Primario ....

Fig. 9-1 — Hoja de respuestas (Test de Coherencia).

Cualquiera sea la forma, en la primera parte del test se le pide al sujeto que estime subjetivamente entre 0% y 100% el porcentaje en que considera existe cada uno de esos cuatro atributos en su universo de discurso. Estos porcentajes se representan como %A - ...%, %B = ...%, %C = ...%, %D =...%. Estos cuatro porcentajes son considerados las **premisas extensionales** del TCR. Estas premisas expresan la idea que el sujeto tiene sobre la extensión o proporción que cada atributo cubre sobre el universo y su existencia o no-existencia; en este último caso se considera al atributo como conjunto o clase vacía.

Por esta razón, los porcentajes — %A, %B, %C y %D — deben ser dados entre 0% y 100% y no tienen por qué sumar 100% todos ellos, sino pueden sumar menos, igual o más que 100%. Este es el porcentaje o extensión del universo. Si alguno de ellos es igual a 100% entonces el atributo correpondiente es tomado como igual al **conjunto universal** (U); y si, por el contrario, alguno de ellos es igual a 0% entonces el atributo

correspondiente es tomado como igual al **conjunto vacío** (0). Estas cotas universales tienen efecto para la evaluación de la coherencia interna del test, ya que si a dos atributos les fueron dados porcentajes que entre ellos suman más del 100%, entonces no pueden mantener entre sí una relación de exclusión en la segunda parte del test porque no pueden exceder la cota

Fig. 9-1 (cont.)

SEGUNDA PARTE

Preg. No.	UNA PERSONA ....(a)....... (atributo sustentado)	TIENE QUE SER	NO PUEDE SER	PUEDE SER	(b) (atrib. interrog.)
1	Una persona AMERICANA	*X*			AMERICANA
2	Una persona AMERICANA	*X*			ARGENTINA
3	Una persona AMERICANA		*X*		EUROPEA
4	Una persona AMERICANA		*X*		INGLESA
5	Una persona ARGENTINA	*X*			AMERICANA
6	Una persona ARGENTINA	*X*			ARGENTINA
7	Una persona ARGENTINA		*X*		EUROPEA
8	Una persona ARGENTINA		*X*		INGLESA
9	Una persona EUROPEA		*X*		AMERICANA
10	Una persona EUROPEA		*X*		ARGENTINA
11	Una persona EUROPEA	*X*			EUROPEA
12	Una persona EUROPEA		*X*		INGLESA
13	Una persona INGLESA		*X*		AMERICANA
14	Una persona INGLESA		*X*		ARGENTINA
15	Una persona INGLESA			*X*	EUROPEA
16	Una persona INGLESA	*X*			INGLESA
17	Una persona AMERICANA		*X*		NO-AMERICANA
18	Una persona AMERICANA		*X*		NO-ARGENTINA
19	Una persona AMERICANA	*X*			NO-EUROPEA
20	Una persona AMERICANA		*X*		NO-INGLESA
21	Una persona ARGENTINA		*X*		NO-AMERICANA
22	Una persona ARGENTINA		*X*		NO-ARGENTINA
23	Una persona ARGENTINA	*X*			NO-EUROPEA
24	Una persona ARGENTINA		*X*		NO-INGLESA
25	Una persona EUROPEA		*X*		NO-AMERICANA
26	Una persona EUROPEA		*X*		NO-ARGENTINA
27	Una persona EUROPEA	*X*			NO-EUROPEA
28	Una persona EUROPEA		*X*		NO-INGLESA
29	Una persona INGLESA		*X*		NO-AMERICANA
30	Una persona INGLESA		*X*		NO-ARGENTINA
31	Una persona INGLESA		*X*		NO-EUROPEA
32	Una persona INGLESA	*X*			NO-INGLESA

(continua)

Fig. 9-1 (cont.)

SEGUNDA PARTE (cont.)

Preg. No.	UNA PERSONA .... (A) (atributo sustentado)	TIENE QUE SER	NO PUEDE SER	PUEDE SER	(B) (atrib. ¡nterrog.)
33	Una persona NO-AMERICANA		X		AMERICANA
34	Una persona NO-AMERICANA		X		ARGENTINA
35	Una persona NO-AMERICANA			X	EUROPEA
36	Una persona NO-AMERICANA			X	INGLESA
37	Una persona NO-ARGENTINA			X	AMERICANA
38	Una persona NO-ARGENTINA		X		ARGENTINA
39	Una persona NO-ARGENTINA			X	EUROPEA
40	Una persona NO-ARGENTINA			X	INGLESA
41	Una persona NO-EUROPEA			X	AMERICANA
42	Una persona NO-EUROPEA			X	ARGENTINA
43	Una persona NO-EUROPEA		X		EUROPEA
44	Una persona NO-EUROPEA			X	INGLESA
45	Una persona NO-INGLESA			X	AMERICANA
46	Una persona NO-INGLESA			X	ARGENTINA
47	Una persona NO-INGLESA			X	EUROPEA
4ª	Una persona NO-INGLESA		X		INGLESA
49	Una persona NO-AMERICANA	X			NO-AMERICANA
50	Una persona NO-AMERICANA		X		NO-ARGENTINA
51	Una persona NO-AMERICANA		X		NO-EUROPEA
52	Una persona NO-AMERICANA			X	NO-INGLESA
53	Una persona NO-ARGENTINA			X	NO-AMERICANA
54	Una persona NO-ARGENTINA	X			NO-ARGENTINA
55	Una persona NO-ARGENTINA			X	NO-EUROPEA
56	Una persona NO-ARGENTINA			X	NO-INGLESA
57	Una persona NO-EUROPEA	X			NO-AMERICANA
58	Una persona NO-EUROPEA			X	NO-ARGENTINA
59	Una persona NO-EUROPEA	X			NO-EUROPEA
60	Una persona NO-EUROPEA			X	NO-INGLESA
61	Una persona NO-INGLESA	X			NO-AMERICANA
62	Una persona NO-INGLESA	X			NO-ARGENTINA
63	Una persona NO-INGLESA			X	NO-EUROPEA
64	Una persona NO-INGLESA	X			NO-INGLESA

universal. Esta restricción está destinada a asegurar la coherencia o consistencia interna entre la extensión y la existencia de los cuatro atributos y las relaciones que guardan entre sí.

A tal efecto, en la segunda parte del test, se le presenta al sujeto una lista de 64 preguntas, en cada una de las cuales debe establecer optativamente una relación entre cada par posible de los cuatro atributos, según

éstos son alternados como antecedente o consecuente de la relación (es decir: como primer término o segundo término del par), y según van siendo afirmados o negados alternativamente. Las negaciones o complementos de A, B, C y D, se representan como -A, -B, -C y -D, respectivamente.

Las respuestas que los sujetos deben dar a cada uno de los items del test (64 de la segunda parte) consisten en tres posibles relaciones, una de las cuales deben optar poner entre los términos de cada par de atributos. Estas relaciones son: 1) **tiene que ser, 2) puede ser, y 3) no puede ser.** Aunque dichas relaciones están expresadas de este modo en la hoja de respuesta, a los efectos de una notación abreviada se representan con los símbolos 1, P y 0, respectivamente. En la hoja de respuestas, el sujeto opta por una de estas tres relaciones marcando una x en la columna correspondiente a una de ellas (al efecto, ver Figura 9-1) y en la línea correspondiente al par de atributos relacionados.

A los efectos de ejemplificar la notación abreviada de las respuestas del sujeto, si éste ha optado por las relaciones "Americano puede ser Argentino" y "Argentino tiene que ser Americano" como atributos A y B, entonces dichas expresiones se notan "APB"y "BIA", respectivamente. De la misma forma, "Americano no puede ser Europeo" se nota "AOC".

Fundamentos lógico-cognitivos del TCR

Las relaciones "tiene que ser", "puede ser" y "no puede ser" las hemos incluido en el TCR como modalidades de la cópula (es, son). Estas modalidades se han arbitrado a los efectos de restarle a la cópula la ambigüedad que reviste en las expresiones silogísticas clásicas — sobre todo en las correspondientes a los enunciados categoriales aristotélicos antes vistos, en donde la comprensión de la cópula está ambiguamente ligada a los términos cuantitativos "todos", "algunos", "ninguno"; por ejemplo, el enunciado universal "Todos los A son B" no excluye la posibilidad de "Algún A es B", y viceversa, dado que dichas proposiciones, en la lógica clásica, son superimplicante y subimplicante, respectivamente, entre sí (Stebbing, 1965).

La antes aludida Henle (1978) destacó esta ambigüedad en la expresión clásica de la cópula, como un factor perturbador en la comprensión correcta de las premisas o del alcance de sus términos. Otros autores como Ceraso y Provitera (1971), que siguieron esas prevenciones de Henle, desarrollaron técnicas silogísticas que modificaban las expresiones clásicas en favor de una menor ambigüedad, y lograron mediante ello que los sujetos alcanzaran mejores niveles de rendimiento lógico. Ello fue atribuído a que, mediante esas modificaciones "desambiguantes", los sujetos alcanzaran una mejor comprensión lógica de las relaciones implícitas en las premisas. Nuestras expresiones modales de la cópula, entonces, antes vistas (representadas por 1, P ó 0), nos han permitido obtener una

mayor comprensión de esas relaciones sin ambigüedad y de una manera más sencilla, por tratarse de expresiones usuales del lenguaje común.

Antes invocamos a ciertos postulados de la lógica modal de Von Wright y entre ellos aludimos a principios modales de **necesariedad,** de **posibilidad y** de **imposibilidad.** Justamente, estos tres principios modales que afectan a la cópula, se hacen efectivos en las expresiones "tiene que ser" (necesariedad), "puede ser" (posibilidad) y "no puede ser" (imposibilidad). Como antes dijéramos, la combinación de principios de una lógica asertiva y de la verdad con principios de una lógica modal como los invocados, parecen afinar o aguzar mucho mejor el sentido y comprensión lógica de la cópula en relación con su uso y alcance real, reduciendo, de este modo, el riesgo o margen de error y ambigüedad en su rol relacionante de términos.

De igual manera, la expresión "A tiene que ser B" ó "AIB" debe entenderse extrictamente como "Todo A es B" a la manera de la lógica proposicional, o bien como "A está incluido en B" en lógica de clases o de conjuntos. Pero a diferencia de la expresión clásica, la expresión "Todo A es B", ó AIB, no es acá compatible ni superimplicante de "Algún A es B". Por la misma razón, la relación dada en la expresióm "A puede ser B" debe entenderse, **desprovista de ambigüedad,** como "Algún A, pero no todos, son B", o "Algún A es B y algún A no es B", proposicionalmente; o como "A tiene intersección no-vacía con B, sin estar incluida en B", en lógica de conjuntos. Esto hace a las relaciones "tiene que ser" y "puede ser" incompatibles entre sí, y portanto inambiguas e inconfundibles. Finalmente, del mismo modo, "A no puede ser B" ó "AOB" debe entenderse como equivalente a "Ningún A es B" o a "A excluido de B", intersección AB vacía, respectivamente. Esta relación también es incompatible con las dos anteriores y con "Algún A no es B", lo cual las hace no-ambiguas entre sí.

Nuestra tesis fundamental aquí, es que el pensamiento, para hacerse explícito, puede crear y aplicar distintas expresiones lógico-formales para representar un mismo sistema de relaciones de la realidad, pero para que esto sea válido debe haber coherencia, transformabilidad, no-ambigüedad y plena diferenciación entre las condiciones específicas de definición de una expresión lógico-formal a otra.

Para cada par de atributos, entonces, el sujeto debe optar por una de esas tres relaciones modales señaladas (1, P ó 0), marcando una cruz en una de ellas para cada item en la columna correspondiente. Para una mejor ilustración, la hoja de respuesta dada en Figura 9-1 se presenta con los porcentajes y marcas que un sujeto supuesto ha contestado en la primera y segunda parte del test.

En el ejemplo de la Figura 9-1 vemos que el "atributo antecedente" (o primer término del par) se presenta fijo para "atributos consecuentes" variables. Esto corresponde a la subforma "a" de la forma II del TCR. La

subforma "b" (o forma lib) supone los mismos atributos pero con "antecedentes variables" y "consecuente fijo". Estas dos subformas "a" y "b" se distinguen para cualquier forma del test y para detectar el posible efecto que pueda tener la variación, o no, del antecedente (o del consecuente) en la coherencia de razonamiento.

Al contestar las primeras doce relaciones entre pares de atributos A, B, C y D, en la segunda parte del test, el sujeto está formulando las **premisas relaciónales.** Estas premisas están constituidas por las relaciones modales 1, P ó 0 que el sujeto marcó (o interpuso) optativamente entre los siguientes pares de atributos: A.B, A.C, A.D, B.A, B.C, B.D, C.A, CB, CD, D.A, D.B y D.C (en la subforma "a"). De cada uno de los seis pares combinados de atributos que pueden construirse a partir de los cuatro atributos originales A, B, C y D — es decir, a partir de los seis pares combinados AB, AC, AD, BC, BD y CD — se desprende una serie de ocho relaciones modales, según que los atributos de cada par vayan siendo alternados como antecedente/consecuente de una relación modal, o bien afirmados o negados. Por ejemplo, a partir del par combinado AB se desprende la siguiente serie de ocho términos:

ArB, BrA, Ar-B, Br-A, -ArB, -BrA, -Ar-B, y -Br-A,

donde r representa genéricamente a una de las tres relaciones modales i, P ó 0; y donde -A y -B son las negaciones o complementos de A y B, respectivamente.

Del mismo modo que hemos desplegado esta serie para el par AB, se pueden desplegar seis series semejantes para los restantes pares combinados; es decir, además de AB, para los pares AC, AD, BC, BD y CD. Estas seis series dan lugar a 4ª items de la segunda parte del test. Si a éstos sumamos las relaciones r que el sujeto debe dar a cada atributo consigo mismo, ya afirmado o negado, se llega entonces al total de 64 items de la segunda parte. Sumados los items de la primera y segunda parte, el TCR queda constituido por 68 items o respuestas consecutivas dadas por el sujeto.

Todo ese conjunto de 68 items constituye en sí una red lógica, dentro de una trama o sistema de relaciones inferenciales. La serie AB arriba desplegada es una serie de inferencias inmediatas. A partir de Stebbing (1965) las hemos calificado como las formas inferenciales: original o directa (ArB), conversa (BrA), obversa (Ar-B), conversa obvertida (Br-A), inversa obvertida (-ArB), contrapuesta (-BrA), inversa (-Ar-B) y contrapuesta obvertida (-Br-A). Estas ocho inferencias inmediatas están lógicamente entrelazadas entre sí, dando lugar a **dependencia lógica dentro de la serle.** Desde ya, esto debe ocurrir también dentro de cada una de las seis series antes mencionadas; pero, dado que estas seis series están inferencialmente entrelazadas entre sí, ello da lugar a una **dependencia lógica entre series,** que debe agregarse a la "dependencia dentro de cada

serie" antes invocada, para lograr la coherencia total del sistema. Tanto la dependencia lógica "dentro" como la "entre" las seis series nos dan una idea de la complejidad y extensión del compromiso lógico que supone para el sujeto el sistema total de inferencias dado por los 68 items del TCR.

Por otra parte, para una única serie de un par de atributos — digamos AB — existen 14 distintas series lógicamente posibles, según la relación que va de A a B sea AIB, AOB ó APB, y según A ó B sean iguales o distintos de U (Universo) y de 0 (vacío). En la Tabla 9-1 se dan estas 14 posibles series AB internamente consistentes. En la columna de la izquierda se dan las igualdades o diferencias entre A, B y las cotas universales U y 0; luego en la columna del medio, se dan las tres posibles relaciones r de A a B como primera premisa relacionai; finalmente en la tercera columna se dan los restantes siete términos de la serie (2.° a 8.º términos) consistentes con lo dado en las columnas primera y del medio de dicha tabla. En la columna del medio las relaciones r (1,0, P) se dan en términos de las relaciones de conjunto caracterizadas por Erickson (1974) en el estudio de razonamiento silogístico. Estas relaciones son:

Id. - Identidad entre A y B (A = B)
Sb. - Subconjunto (A incluida en B)
Sp. - Superconjunto (A incluye a B)
Ov. - Solapado *(overlap)* (A y B se solapan pero no se incluyen)
Ex. - Exclusión (A y B se excluyen mutuamente)
(Sp.) - A Superconjunto cuando B = 0 (conjunto vacío)

Estas mismas 14 series lógicamente posibles pueden darse para los restantes pares AC, AD, BC, BD y CD. Erickson es un investigador del razonamiento deductivo que, en tal sentido, ha seguido supuestos similares a los de Henle.

Sobre la base de todas estas definiciones y compromisos lógicos dados como requisitos para la evaluación del TCR, consideramos probada su pertinencia y validez como instrumento para evaluar y medir la coherencia y organización lógica del pensamiento, sobre la base de la **isomorfía** entre lógica y razonamiento del sujeto.

En tal sentido, la medida de esa coherencia y organización va a depender, y va a estar validada al mismo tiempo, por ese **grado de isomorfismo.** La isomorfía entre la estructura inferencial, secuencial, de las respuestas dadas por el sujeto y la estructura inferencial que se desprende de aplicar principios y reglas lógicas a partir de premisas consistentes dadas por el propio sujeto no puede ser aleatoria ni inatinente.

Como ya hemos sugerido, para medir ese isomorfismo hemos adoptado el criterio de contar los errores de contradicción cometidos por el sujeto a lo largo del TCR. Esos errores de contradicción surgen de aquellos items en los que la relación r dada por el sujeto no corresponde con la que **debe ser** de acuerdo con los principios y reglas lógicas. Estos

Tabla 9-1 — Consistencia dentro de la Serie (AB)

Igualdad/Diferencia entre A, B, U, y 0		Relaciones de conjunto dadas por Erickson — Primera premisa relacionai			Series (AB) posibles consistentes (desde el 2 o al 8.0 término)
		(A1B)	(AOB)	(APB)	
A=U, B=U	(A=B)	Id.		Sp.	(B1A)(A0-B)(B0-A)(-A1B)(-B1A)(-A1-B)(-B1-A)
A=∅, B=U	(A≠B)	Sb.	(Sp)	Sp.	(BOA) (A1 -B) (B1 -A) (-A1B) (-B1 A) (-A0-B) (-B1 -A)
U≠A≠∅, B=U	(A≠B)	Sb.		Ov.	(BPA) (AO-B) (BP-A) (-A1B) (-B1 A) (-A0-B) (-B1 -A)
A=U, B=∅	(A≠B)	Id.	(Sp-)		(B1A)(A1-B)(B1-A)(-A1B)(-B1A)(-A1-B)(-B0-A) (B1
A=∅, B≠∅	(A=B)	Sb,	Ex.		A) (A1 -B) (B1 -A) (-A0B) (-BOA) (-A1 -B) (-B1 -A)
U≠A≠∅, B=∅	(A≠B)	Id.	Ex.		(B1A)(A1-B)(B1-A)(-A0B)(-BPA)(-A1-B)(-BP-A)
A=U, U≠B≠∅	(A≠B)	Sb.			(B1A)(AP-B)(B0-A)(-A1B)(-B1A)(-A1-B)(-B0-A)
A=∅, U≠B≠∅	(A≠B)				(BOA) (A1 -B) (B1 -A) (-APB) (-BOA) (-AP-B) (-B1 -A)
U≠A≠∅, U≠B≠∅	(A=B)				(B1 A)(A0-B) (BO-A) (-A0B) (-BOA) (-A1 -B) (-B1 -A)
A⊂B, B⊄A	(AB≠∅)				(BPA) (AO-B) (BP-A) (-APB) (-BOA) (-AP-B) (-B1 -A) (B1
A⊄B, B⊂A	(AB≠∅)				A) (AP-B) (BO-A) (-A0B) (-BPA) (-A1 B) (-BP-A)
A⊄B, B⊄A	(AB≠∅)				(BPA)(AP-B)(BP-A)(-APB)(-BPA)(-AP-B)(-BP-A)
A⊂-B, -B⊄A	(AB=∅)				(B0A)(A1 -B) (B1 -A) (-APB) (-BPA) (-AP-B) (-BP-A)
A⊂-B, -B⊂A	(AB=∅)				(B0A)(A1-B)(B1-A)(-A1B)(-B1A)(-A0-B)(-B0-A)

Id. =identidad, Sb. =subconjunto, Sp. = superconjunlo, Ov. =Iraslapamiento (overlap), Ex. =exclusión entre conjuntos. Las relaciones Sp. entre paréntesis pueden dar lugar a controversia con respecto a Ex.; ellas respuenden al criterio del autor que se ha basado en (B1 A). Esta ambigüedad surge del conjunto vacio cuando se lo emplea como término consecuente de las relaciones R.

principios y reglas parten de las definiciones preposicionales y de conjuntos que antes viéramos en relación con las relaciones **tiene que ser** (1), **puede ser (P) y no puede ser** (0).

Los errores de contradicción en el TCR son en realidad **autocontradicciones** que el sujeto comete con sus propias premisas, lo que destaca más su carácter interno y de organización del pensamiento. En trabajos anteriores (López Alonso, 1988; Ricardi, 1989) hemos probado que los sujetos cometen menos errores de contradicción cuando son ellos mismos los que eligen sus premisas que cuando se las impone el experimentador "desde afuera". Esta diferencia hablaría ya de por sí a favor de un factor interno de autoconsistencia o autocoherencía, que podría resumirse como que: "todo sujeto tiende a ser más coherente consigo mismo que con el resto del mundo". Sí bien, esto puede entenderse a primera vista como un principio un tanto obvio o trivial, su importancia es fundamental: primero, porque diferencia una medida de la coherencia interna del sujeto de una externa, y, segundo, porque por otros resultados obtenidos parece ser que la coherencia interna es un prerrequisito y condición indispensable para el logro de la coherencia externa o con el mundo (López Alonso, 1991c, 1992; Ricardi, 1993a; López Alonso, Ricardi, 1993).

Los fundamentos lógicos del TCR serán completados en las secciones siguientes referidas a su procedimiento de evaluación y su evaluación mediante diagrama rectangular.

*Procedimiento de evaluación del **TCR** — hoja de evaluación*

Una vez que el sujeto ha completado los items del TCR sus respuestas son volcadas a una hoja de evaluación. Un modelo de esta hoja se da en la Figura 9-2.

PREMISAS EXTENSIONALES

Matriz I

		A	B	C	D
60%	A	1 / 1	///1/// / P	0 / 0	0 / 0
55%	B	1 / 1	1 / 1	0 / 0	0 / 0
30%	C	0 / 0	0 / 0	1 / 1	///0/// / P
25%	D	0 / 0	0 / 0	P / P	1 / 1

Matriz II

	-A	-B	-C	-D
A	0 / 0	///0/// / P	1 / 1	0 / 1
B	0 / 0	0 / 0	1 / 1	0 / 0
C	0 / 1	0 / 1	1 / 0	0 / P
D	0 / 1	0 / 1	0 / P	1 / 0

Matriz III

		A	B	C	D
40%	-A	0 / 0	0 / 0	P / P	P / P
45%	-B	P / P	0 / 0	P / P	P / P
70%	-C	P / P	P / P	0 / 0	P / P
75%	-D	P / P	P / P	P / P	0 / 0

Matriz IV

	-A	-B	-C	-D
-A	1 / 1	0 / 1	0 / P	P / P
-B	P / P	1 / 1	P / P	P / P
-C	1 / P	P / P	1 / 1	P / P
-D	1 / P	1 / P	P / P	1 / 1

Diagrama Rectangular

(+) A-B-C-D	A		(+) -A-B-C-D	U
	(+) AB-C-D	(-) B -AB-C-D		
(-) A-BC-D	(-) ABC-D	(-) -ABC-D	(+) C -A-BC-D	
(-) A-BCD	(-) ABCD	(-) -ABCD	(+) -A-BCD	D
(-) A-B-CD	(-) AB-CD	(-) -AB-CD	-A-B-CD	

ERRORES DE INCONSISTENCIA

Incompatibilidades con:	PUNTAJE
Premisas extensionales	3
Premisas relacionales	13
Identidad: + ó -	-
No-contradicción / 3.º Excl.	2
Preguntas no respondidas	-
TOTAL	18

Fig. 9-2 — Hoja de evaluación.

En la parte de arriba de la hoja de evaluación pueden observarse cuatro matrices de 4x4, subdenominadas matrices I, II, III, IV, en las cuales cada hilera y cada columna corresponden respectivamente a los atributos A, B, C, D; o bien a sus negaciones: -A, -B, -C, -D. En todas estas matrices las hileras corresponden al "atributo antecedente" o primer término de cada par, en tanto que las columnas siempre corresponden a los "atributos consecuentes" o segundo término del par.

Las celdas de las matrices están destinadas al registro de las relaciones modales r que ligan a los correspondientes atributos — el que define a la hilera y el que define a la columna — de las cuales la celda es intersección. Por ejemplo, si un sujeto en la hoja de respuesta respondió "Una persona Americana **puede ser** Argentina" (es decir, marcó x en la columna "puede ser" entre los atributos Americano y Argentino), esto se transcribe brevemente como APB y se registra P en la celda intersección de la hilera A y la columna B. Asimismo, si este sujeto luego responde "Una persona Argentina **tiene que ser** Americana", o B1A, entonces se coloca 1 en la celda intersección de la hilera B y la columna A.

Las celdas están subdivididas por una línea de puntos en una parte superior y una inferior. En la parte superior se anota la relación r que el sujeto dio al par de atributos correspondientes. La parte de abajo, en cambio, está destinada al evaluador, para que éste anote allí la relación r que corresponde de acuerdo con principios lógicos. **Si ambas r coinciden** habrá ¡somorfía entre la estructura de las relaciones del sujeto y las de la lógica; pero **si no coinciden** se marcará un determinado error **de** contradicción.

En la matriz I se anotan las relaciones r dadas a los pares de atributos afirmados; por ejemplo ArB y BrA. En esta matriz van las inferencias originales y conversas. En la matriz II van las relaciones entre antecedente-afirmado y consecuente-negado; por ejemplo, Ar-B y Br-A. Siguiendo a Stebbing, en esta matriz van las inferencias obversas y conversas obvertidas. En la matriz III van las relaciones antecedente-negado/consecuente-afirmado — por ejemplo, -ArB y -BrA — y corresponde a las inferencias inversa obvertida y contrapuesta. Y finalmente, en la matriz IV van las relaciones antecente y consecuente ambos negados — por ejemplo, -Ar-B, -Br-A — correspondiendo a las inferencias inversa y contrapuesta invertida.

En cada una de estas matrices, la diagonal que va desde el ángulo superior-izquierdo al ángulo inferior-derecho contiene las relaciones que pueden establecerse entre afirmaciones y/o negaciones de un mismo atributo. Así, por ejemplo, en la matriz I está la relación entre A y A (AIA), etc.; en la matriz II, la relación entre A y -A (AO-A), etc.; en la matriz III, -A y A (-A0A), etc.; y en la matriz IV, la relación entre -A y -A (-AI-A), etc. Las relaciones de las diagonales de las matrices I y IV son de

identidad positiva y de **identidad negativa,** respectivamente, y se les da la relación 1 (tiene que ser). Las relaciones de las diagonales de las matrices II y III registran relaciones básicas de **no-contradicción** y de **tercero excluido,** y se les da la relación 0 (no puede ser).

A la izquierda de la matriz I, se anotan los porcentajes correspondientes a las premisas extensionales %A, %B, %C y %D. Estos valores controlan que las premisas relaciónales no sean contradictorias con las premisas extensionales y las cotas universales del sistema. En caso de que exista alguna contradicción entre ambos tipos de premisas — por ejemplo: que el sujeto haya contestado %A = 50% y %B = 60%, y que después pretenda establecer las relaciones B1A (B incluido en A) o bien AOB (A y B excluidos entre sí) — las mismas se computan como **errores extensionales,** dado que constituyen incompatibilidades con las premisas extensionales. Las premisas relaciónales, además de deber ser compatibles con las extensionales, deben ser compatibles entre sí, y asimismo todas las restantes relaciones r en el test deben ser compatibles con ambos tipos de premisas.

Por debajo de las cuatro matrices antes vistas, se da un **diagrama rectangular** (inspirado en diagramas de Venn) a través del cual se logra un **mapa de la estructura de relaciones modales correspondientes a las premisas relaciónales compatibles dadas por el sujeto.** Con este mapa el evaluador decide la compatibilidad que las premisas relaciónales deben guardar con las premisas extensionales y luego las relaciónales entre sí, como así también la compatibilidad que las restantes respuestas del test deben guardar con todas las premisas.

A los efectos de ilustrar el procedimiento, en la hoja de evaluación de la Figura 9-2 se vuelcan los datos anotados en la hoja de respuesta de la Figura 9-1.

Evaluación mediante el diagrama rectangular

La evaluación del TCR se realiza completamente a través del diagrama rectangular que aparece en el ángulo inferior izquierdo de la Figura 9-2. Este diagrama aparece subdividido en 16 espacios como consecuencia de superponer en el mismo rectángulos correspondientes a los atributos A, B, C y D. De este modo, el rectángulo mayor corresponde al conjunto universal (U), y los rectángulos menores superpuestos corresponden a dichos cuatro atributos. Para reconocer en el gráfico cuál es cada uno de esos rectángulos, se han colocado las letras identificatorias correspondientes (U, A, B, C y D) en los ángulos superior-derecho de cada uno.

A los efectos de la evaluación, el evaluador debe llenar los 16 espacios con signos + y -. El signo + indica que un espacio o intersección es no-vacío; en tanto que el signo - indica que el espacio es vacío. La distribución de estos signos en los 16 espacios del diagrama depende

exclusivamente de las relaciones que el sujeto dio a las doce premisas relaciónales. Los signos +/- varían según la relación r que el sujeto estableció entre dos atributos.

Para una mejor ejemplificación de las reglas con que se asignan los signos +/-, se dan en la Figura 9-3 diagramas de Venn explicativos. En ellos, se da la distribución de signos +/- para las relaciones 1, P y 0 entre los atributos A y B, cuando A es antecedente y B consecuente. Los cuatro casos que allí se plantean deben entenderse como reglas lógicas de evaluación según que el atributo antecedente (en este caso A) sea igual o distinto de vacío. El signo ? que aparece en el espacio de B, fuera de A, es para ser llenado por la conversa BrA. Con estas reglas básicas se llena todo el conjunto de 16 espacios del diagrama rectangular, siguiendo el siguiente criterio: **Cada vez que una relación r obliga a poner en un espacio un signo +/- distinto del que ya había sido puesto por otra relación r anterior, esta discrepancia indica que estamos frente a una contradicción.** En tanto los signos coincidan habrá coherencia o compatibilidad — y también **Isomorfismo** lógica/razonamiento del sujeto.
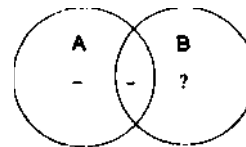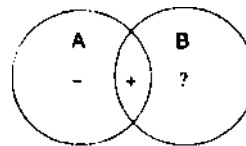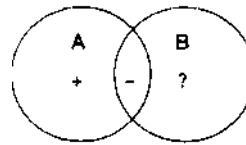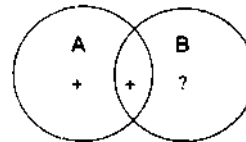
Fig. 9-3 — Diagramas de Venn.

Cómo se establece el puntaje del sujeto en el TCR

El puntaje que un sujeto obtiene en el TCR se basa en el cómputo de respuestas erróneas (o contradicciones) y no en el de respuestas correctas. De este modo, un sujeto que obtiene 0 puntos en el test es un sujeto de coherencia perfecta, y si su puntaje aumenta, ello será un indicador de un grado progresivo de incoherencia de razonamiento o de insuficiencia de su organización lógica del pensamiento.

El cifrar el puntaje sobre la base de los errores y no en las respuestas correctas tiene una serie de ventajas. La primera de ellas es la posibilidad de diferenciar distintos tipos de errores y una segunda es analizarlos en profundidad posteriormente, según un análisis de items correspondientes. Otra ventaja sería detectar las posibles falencias estructurales del razonamiento del sujeto en sus puntos precisos de fracaso inferencial.

Pero el aspecto más importante por el cuál conviene centrar el cómputo más en los errores que en las respuestas correctas, es que esto le permite al investigador pasar de un enfoque evaluativo estrecho y superficial, basado en la respuesta (modelo estímulo-respuesta) a un enfoque más amplio y profundo basado en los procesos inferenciales puestos en juego a lo largo del test (modelo estímulo-proceso inferencial ¡ntraorganísmico-repuesta). De este modo, esto permite estudiar las causales de incoherencia y de desorganización lógica del pensamiento. Esto se ha logrado en trabajos sucesivos (López Alonso, 1988, 1991a, 1991b, 1992; Orsi, 1988; Ricardi, 1989, 1993a, 1993b) y se continuará.

El puntaje del test se basa entonces en el cómputo de errores de contradicción cometidos por el sujeto, los cuales son discriminados según los siguientes tipos:

Contradicciones con las premisas extensionales

Se trata de aquellas relaciones r que entran en contradicción con los porcentajes %A, %B, %C y %D. Se computa la cantidad de errores de este tipo.

Contradicciones con las premisas relaciónales

Son las relaciones r que entran en contradicción con otras relaciones r previamente dadas o correspondientes a las premisas relaciónales. También se computa separadamente la cantidad de errores de este tipo.

Transgresiones al principio básico de identidad

Se trata de aquellas relaciones r incompatibles entre dos afirmaciones o dos negaciones de un mismo atributo. Este tipo de error aparece en las diagonales antes señaladas de las matrices I y IV. Se distinguen dos subtipos: 1) Transgresiones a la identidad positiva (ó Id.+) — son los

errores de contestar, por ejemplo, AOA ó APA, en lugar de AIA, que es la identidad. Estos errores sólo aparecen en la matriz I y se computan como tales; 2) Transgresiones a la identidad negativa (ó Id.-) — son los errores de contestar, por ejemplo, -AO-A ó -AP-A en lugar de -AI-A , que es la identidad de un atributo definido por lo negativo. Este tipo de identidad es importante distinguirlo, porque se ha dado como error significativo en ciertos tipos de cuadros o perturbaciones psiquiátricas (López Alonso, 1988). Estos errores sólo aparecen en la matriz IV y se computan como tales.

Transgresiones a principios básicos de no-contradicción y tercero excluido

Se trata de aquellas relaciones r incompatibles dadas entre la afirmación y la negación de un mismo atributo. Por ejemplo, contestar AI-A ó AP-A, en la diagonal de la matriz II; o -AIA ó -APA en la de la matriz III, en lugar de AO-A ó -AOA, respectivamente. Estos errores pueden computarse conjunta o separadamente, pues ambos transgreden una noción básica de no-contradicción y de tercero excluido, según se vea, y son un índice indudable de desorganización lógica.

Items del TCR no respondidos

Se trata de las relaciones r que el sujeto ha dejado en blanco. Estas no-respuestas se consideran como una forma de fracaso en el logro de la coherencia total, aunque no constituyan en sí algún tipo de contradicción. Por tal razón, se computan como tales, separadamente.

Puntaje total de errores

Bajo este título se da la suma total de errores o fracasos dados en los cinco puntos anteriores. Esta suma total es un reflejo del rendimiento global del sujeto en coherencia de razonamiento.

El cómputo de estos seis puntos se da en el ángulo inferior-derecho de la Figura 9-2 y con el mismo culmina y se sintetiza el proceso de evaluación llevado a cabo en esa hoja. Recuérdese que cuanto más alto es el puntaje o número de errores, mayor es el índice de incoherencia y desorganización lógica del pensamiento. Luego, para un análisis cualitativo también importa tener en cuenta el tipo de error, su ubicación dentro de la cadena inferencial y el tipo de inferencia inmediata (según Stebbing) que supone.

Formas del TCR desarrolladas hasta el presente

Como puede apreciarse a través de la lectura de los puntos anteriores, el TCR tiene un esquema básico que parte de un conjunto de atributos y que puede variarse indefinidamente según cuáles sean los atributos que se tomen y cuál es el área de aplicación o investigación científica.

Ya hemos hecho referencia a las Formas I y II del test y a sus subformas a y b correspondientes. Recordemos que la Forma I estaba compuesta por los atributos "Respetable", "Honorable", "Admirable" y "Venerable", que son formas incluidas de valoración social y cultural. Recordemos también que la Forma II estaba constituida a partir de los atributos "Americano", "Argentino", "Europeo" e "Inglés", que pueden interpretarse como nacionalidades o lugares de nacimiento.

A estas dos formas ya referidas, se le agregaron posteriormente las siguientes formas:

Forma III — constituida por los atributos "Americano" (continental), "Argentino", "Hispanoparlante" y "Angloparlante". Esta forma se caracteriza por combinar dos atributos de la Forma II con otros dos atributos lingüísticos, dando lugar a una estructura relacionai distinta.

Forma IV — desarrollada para niños en colaboración con la Lio Nora Ricardi. Los atributos constitutivos fueron: "Los animales", "Los gatos", "Las flores" y "Las rosas", y se presentaban en forma textual y en forma gráfica simultáneamente, a efectos de facilitar la comprensión de los niños. Esta forma fue ideada especialmente para ser comparada con pruebas de desarrollo intelectual de Piaget, tanto en niños normales, como diferencia-damente en niños oyentes y en niños sordos (Ricardi, 1989, 1993b) repectivamente.

Forma V — constituye una ampliación metodológica e instrumental significativa del TCR, pues consiste en extender la evaluación de la coherencia inferencial desde el conjunto básico de cuatro atributos a un conjunto de atributos mucho mayor. Por ejemplo, el primer modelo de la Forma V se constituyó sobre los cuatro atributos básicos, que en este caso estuvieron referidos a posturas y categorías de valor como: "Idealista", "Realista", "Materialista" y "Escéptico". Además de establecerse las relaciones r entre estos cuatro atributos y su correspondiente diagrama rectangular, se establecieron las relaciones r que guardaban esas cuatro posturas de valor con cada uno de los términos de una lista de 36 actividades u ocupaciones humanas. Entre éstas estaban, por ejemplo: "sacerdote", "científico", "político", "prostituta", "juez", "narcotraficante", "médico", "educador", "travesti", "economista", "banquero", "drogadicto", "asistente social", "psicólogo", "abogado", "predicador", "legista", "filántropo", etc.

Esta nueva forma pemítió establecer un mapa para cada una de esas actividades o profesiones según su compatibilidad o incompatibilidad con las relaciones del diagrama rectangular que refleja las posturas axiológicas del respondente y dando el grado de "idealismo", "realismo", "materialismo" y "escepticismo" que corresponda a cada una de esas 36 actividades. Sólo a modo de breve referencia, diremos que el mapa de "sacerdote" se constituyó con mayor frecuencia y consistencia sobre las

intersecciones de los atributos "idealista" y "realista"; mientras que el de "narcotraficante" se constituyó con igual definición sobre las intersecciones de "materialista" y "escéptico". De este modo, tomando estos dos ejemplos contrapuestos, puede establecerse y ampliarse indefinidamente el mapa inferencial que liga a distintos atributos a través de distintas categorías conceptuales importantes, y a distintos niveles de significación. Dichos mapas constituyen en sí una estructura de relaciones inferenciales entre conceptos que permiten explorar, justamente, procesos de formación de conceptos, de categorización, definición y aprendizaje. Las posibilidades de exploración, de este modo, se tornan ilimitadas, dado que sólo depende de cambiar el conjunto de términos o atributos que se desee estudiar. Una de las aplicaciones posibles que se está planeando para un futuro — esperemos — no muy lejano, es la de utilizar esta forma para establecer el grado de consistencia o no-contradicción **entre y dentro** distintos **sistemas expertos** referidos a áreas comparables de conocimiento, donde a veces la no-consistencia o contradicción mutua entre y dentro de los sistemas expertos haya creado gran desconcierto. Obviamente, esta aplicación supone crear formas del TCR especializadas y ajustadas a cada sistema experto.

Por último, otra forma que se proyecta es una en la cuál se combinan percepción visual de la realidad, organización perceptual, memoria operativa, coherencia de razonamiento y organización lógica del pensamiento.

CONCLUSIONES

Hemos reservado para esta última sección las consideraciones relativas a las normas psicométricas del TCR y a algunos de los resultados obtenidos con dicho instrumento.

Respecto de las normativas psicométricas esperadas de todo instrumento de medición, en especial a las propiedades de confiabilidad, validez, sensibilidad y generabilidad que deben satisfacer dichos instrumentos, cabe decir lo siguiente:

La primer propiedad que hemos atendido y — entendemos — satisfecho es la de **validez.** Creemos que por sus fundamentos lógicos, el TCR tiene de por sí una validez de contenido y de constructo asegurada en esa disciplina. Secundariamente y un poco contrariamente a lo habitualmente esperado, hemos puesto atención en la **confiabilidad.**

Nuestro criterio extemo ha sido en realidad el de la i**somorfía** entre los procesos de razonamiento del sujeto y la lógica, y entendemos con esto asegurada la validez de la prueba sobre la base de que esa isomorfía no es un producto aleatorio, sino lógico-organizacional de la mente. La

posibilidad de que esa isomorfía o coincidencia estructural entre dos sistemas independientes (lógica-pensamiento) ambos tan complejos, se produzca al azar es totalmente impensable. Las posibilidades aleatorias de llenar al azar sólo la segunda parte de la hoja de respuestas es de 3 a la 64. Por otra parte, llama la atención la habitual frecuencia (30%) con que jóvenes sin instrucción lógica son capaces de responder el TCR sin ningún error de contradicción, la que es mucho más alta cuando se trata de uno a cinco errores solamente. Todo esto ha garantizado de algún modo no sólo la validez del instrumento, sino la existencia de una organización lógica natural en la mente humana (López Alonso, 1990, 1991a) y éste es uno de los resultados más importantes.

Implícitamente, nuestro criterio de validez se ha sustentado también en el principio wundtiano de heterogénesis de los fines, dado que, adaptando nuestra interpretación del mismo, la lógica se instrumenta como una disciplina independiente del pensamiento, aunque se haya originado a partir de éste. En tal sentido, si la lógica no fuera independiente del pensamiento o si no supusiéramos dicho principio, y no fuera, por tanto, un sistema regulativo y evaluativo del mismo independiente, entonces no podríamos adoptar la lógica como "criterio externo" para fundamentar sobre el mismo nuestro criterio de validez.

Volviendo a la confiabilidad, como antes dijéramos, la hemos dejado en un plano secundario. Con esto hemos seguido un criterio que ha sido destacado por Sechrest (1989). Este autor ha objetado críticamente a aquéllos que, sin pensarlo, suponen que una medida no puede ser válida sin ser previamente confiable. Para ello, Sechrest se ha basado en el criterio de seguridad y generalización, previos al de confiabilidad, tal como ha sido sustentado por Cronbach y otros (1972), y también se ha basado en ejemplos críticos propios y de Harrell (1981) respecto de la estabilidad y seguridad de las medidas de confiabilidad.

Según estos autores, si bien la confiabilidad es usualmente interpretada como un coeficiente que indica el grado de estabilidad o no-variabilidad aleatoria de una medición, puede no diferenciar suficientemente entre la estabilididad/variabilidad del instrumento y la estabilidad/variabilidad del objeto o sujeto medido. Para evitar este riesgo, debemos contar con un suficiente criterio de validez previo.

Si tenemos asegurado un criterio de validez tal, se atenuará esa insuficiencia de la confiabilidad frente a la complejidad de factores del proceso medido. En tal sentido, la confiabilidad no ocultará riesgo o error alguno de estimación y/o evaluación por variabilidad no discriminada. Sechrest atribuye también a falta de comprensión del problema de la generalización visto por Cronbach y colaboradores y a la confusión derivada de distintos conceptos de confiabilidad, como una fuente posible de graves errores en la fundamentación de la medición. Siguiendo este

criterio, hemos preferido, entonces, asegurar previamente un suficiente criterio de validez sustentado en el isomorfismo entre la lógica deductiva de clases y/o de conjuntos y el razonamiento natural o no entrenado lógicamente de los sujetos.

No obstante, hemos obtenido medidas de confiabilidad "test-retest" y "formas paralelas" bastante satisfactorias — correlaciones iguales a 0,95 y 0,89, respectivamente.

Respecto a la sensibilidad y generabilidad del TCR, cabe decir que éstas dependen del mismo factor internológico-organizacional que se pretende medir. López Alonso (1992), Ricardi (1993a) y López Alonso y Ricardi (1993) constataron, en los resultados del TCR tomado a muestras de estudiantes secundarios, de sujetos psiquiátricos y controles, que la sensibilidad básica del instrumento y el grado de organización lógica del pensamiento están fuertemente asociados y dependen de la estructura de relaciones r que los sujetos han dado como premisas relaciónales.

Asimismo, respecto a la generalidad, constataron en los mismos trabajos que ésta aumenta cuando la estructura de relaciones r que el sujeto confiere a las premisas se adecúan mucho más a las relaciones existentes en la realidad. Los autores mencionados diferenciaron para una misma forma del TCR (Forma II) 40 estructuras diferentes de relaciones de premisas. Una sola de esas estructuras acaparó la mayor frecuencia o adhesión por parte de los sujetos; esta estructura fue la que más unívocamente reflejó la estructura de relaciones de la realidad y la que dio el promedio más bajo de errores de contradicción (3,90 errores para 176 sujetos), comparada con estructuras unipersonales que distorsionaban las relaciones de la realidad y donde el número de errores ascendió a 57 sobre 64 ítems (para un único sujeto).

Como vemos, el TCR refleja su sensibilidad através del distinto tipo de estructura relacional de premisas dada por los sujetos, y refleja su generalidad a través del grado en que dicha estructura se ajusta de la forma más adecuada y unívoca a la estructura de relaciones existentes en la realidad, y no la distorsiona. Con estos resultados se comprueba cómo el TCR satisface y pone de relieve la importancia de la diferenciación antes invocada de que todo sujeto tiene dos formas de coherencia: una, interna, consigo mismo, y otra, externa, con el mundo y la realidad — donde la primera importa para el logro de la segunda, pues la primera representa a su organización lógica interna.

Respecto a su fundamentación epistemológica, el TCR es una técnica coincidente con los principios y estructuras cognitivas fundamentales de la "lógica operatoria" de Piaget (1977), ya sea en lo que hace a los aspectos de extensión y comprehensión de las estructuras formales jerárqui-

cas operativas de agrupamiento, grupo y/o reticulado, como a sus aspectos de conservación, coordinación, reversibilidad y equilibrio del sistema. Y coincide también con Piaget en el criterio de considerar que la capacidad de resolución de las contradicciones se halla en el nivel más alto y avanzado del desarrollo intelectivo.

Queda una última cuestión: Hemos supuesto a la lógica como independiente del pensamiento, pero ¿hasta qué punto el grado de isomorfismo entre ambos es una medida cabal de la organización del pensamiento y de su dinámica?

Ya hemos dicho que el pensamiento es un proceso vital y abierto, y por lo tanto expuesto permanentemente a nuevos conflictos cognitivos y contradicciones, lo cual lo hace cambiante, sujeto a error, incertidumbre e inestabilidad. Esto es producto de su intercambio dinámico con la realidad y depende de las representaciones, intencionalidad, objetivos y expectativas del sujeto pensante. En otro trabajo (López Alonso, 1990, 1991a) nos referimos a esta característica del pensamiento como sistema abierto y necesitado de un permanente equilibrio entre redundancia y entropía de la mente.

Cuando iniciamos nuestro proyecto de medir la organización lógica del pensamiento a través del TCR, no partimos de la idea de que el pensamiento tuviera que ser una organización lógica perfecta e infalible. Una organización tal nos estaría hablando de un sistema cerrado y autosuficiente, cuando, en realidad, el pensamiento es todo lo contrario: un sistema abierto, falible y perfectible, permanentemente expuesto a error e incertidumbre. Si el pensamiento no fuera así, no sería ni pertenecería a un sistema viviente.

Pero entonces, dentro de ese contexto ¿qué sentido tiene medir la organización lógica del pensamiento a través de su isomorfismo con la lógica — una ciencia formal, de "artefactos formales congelados", atemporal e inamovible?

Nuestra respuesta es: El isomorfismo, en realidad, está midiendo la coherencia y la fuerza lógica del pensamiento hasta un cierto estándard o nivel de alcance y realización representado por una cierta lógica, y no más allá de ésta. La lógica sólo se encarga de dar y marcar ese nivel. Es como una "regla" con la que se mide el crecimiento organizacional del pensamiento y sólo nos dice que el pensamiento ha llegado "por lo menos hasta acá" en su organización, pero esa "regla" no es esa organización ni la explica. El pensamiento puede crecer en su complejidad mucho más allá de lo que esa "regla" o lógica puede medir, pues el pensamiento es un proceso vital, inferencialmente progresivo y sin límites — un generador ilimitado de significaciones, conceptos y formas.

ABELSON, R.P. ARONSON, E., MCGUIRE, W.J.. NEWCOMB, T.M., ROSENBERG, M.J.. TANNEMBAUM, P.H. (Eds.)- *Theories of cognitive consistency,* a source-book. Chicago (IL): Rand-McNally, 1968.

BENTHAM, J. *Outline on a new system of logic.* [S.I.: s.n.], 1827.

BOOLE, G. *An investigation of the laws of thought.* London: Macmillan, 1854.

--------------- .*The mathematical analysis of logic.* Cambridge (MA): MacMillan, Barclay and MacMillan, 1847.

--------------- . *The mathematical analysis of logic.* Oxford: Basil Blackwell, 1948,1951.

BROUWER, L.E.J. Intuitionism and formalism. *Bulletin of the American Mathematical Society,* V.20, p.81-96, 1913.

CERASO, J., PROVITERA. A. Sources of error in syllogistic reasoning. *Cognitive Psychology.* V.2, p.400-410. 1971.

CHAPMAN. L.J., CHAPMAN, J.P. Atmosphery effect re-examined. *Journal of Experimental Psychology* Washington, D.C.. v.58, p.220-226,1959.

CRONBACH. L.J., GLESER, G.C., RAJARATNAM, N.. NANDA. H. *The dependability of behavioral measure,* theory of generalizability for scores and profiles. New York: Wiley, 1972.

DEL VAL, J.A. (Comp.). *Investigaciones sobre lógica y psicología.* Madrid: Alianza Universidad, 1977.

DICKSTEIN, L.S. Differential difficulty of categorical syllogisms. *Bulletin of the Psychonomic Society.* V.8, p.330-332. 1976.

ERICKSON. J.R. Research on syllogistic reasoning. In: REVLIN, R., MAYER, RE. (Eds.). *Human reasoning.* Washington, DC: Winston, 1978.

--------------- . A set analysis theory of behavior in formal syllogistic reasoning tasks. In: SOLSO, R.L. (Ed.). *Theory of cognitive psychology,* the Loyola Symposium. Potomac: Erlbaum. 1974.

EVANS, J.S.B.T. Selective processes in reasoning. In: EVANS, J.S.B.T. (Ed.). *Thinking and reasoning,* psychological approaches. London: Routledge and Kegan. 1983.

FEATHER, NT. Acceptance and rejection of arguments in relation to attitude strength, critical ability, and intolerance of incosistency. *Journal of Abnormal and Social Psychology,* v.69, p. 127-136,1965.

FELDMAN, S. (Ed.). *Cognitive consistency,* motivational antecedents and behavioral consequents. New York: Academic Press, 1966.

FESTINGER, L *A theory of cognitive dissonance.* Stanford (CA): Stanford University Press, 1957.

FRANDSEN, A.N., HOLDER. J.R. Spatial visualization in solving complex verbal problems. *Journal of Psychology.* Provincetown (MA), v.73, p.229-233, 1969.

FRASE, L.T. Belief, incongruity, and syllogistic reasoning. *Psychology Reports,* v. 18, p.982, 1966.

GOTTESMAN, L. CHAPMAN, L.J. Syllogistic reasoning errors in schizophrenia. *Journal of Consulting Psychology,* v.24, p.250-255, 1960.

GUILFORD. J.P. Three faces of intellect. *American Psychologist.* Washington. D.C., v.14, p.469-479, 1959.

HALPERN. J. Reasoning. In: EYSENCK. M.W. (Ed.). *The Blackwell dictionary of cognitive psychology.* Oxford: Basil Blackwell. 1990.

HAMILTON, W. *Lectures on metaphysics and logic.* [S.I.: s.n.J, 1859.

HARRELL. W.A. The effects of alcohol use and offender remorsefulness on sentency decisions. *Journal of Applied Social Psychology.* Silver Spring (MD). v. 11, p.83-91, 1981.

HEIDER, F. *The psychology of interpersonal relations.* New York: Wiley. 1958.

HENLE. M. Foreword. In: REVLIN, R., MAYER. R.E. (Eds.). *Human reasoning.* Washington, D.C.: Winston. 1978.

. ------------ . On the relation between logic and thinking. *Psychological Review.* Washington. D.C.. V.69, p.366-378, 1962.

HERBART, J.F. *Hauptpunkte derlogik.* [S.I.: s.n). 1808.

— ---------- . *Lehrbuch zur einleitung in die philosophie.* [S.I.: s.n.]. 1813.

JOHNSON-LAIRD, P.N. *Mental models.* Cambridge (MA): Harvard University Press. 1983.

------------- . *El ordenador y la mente:* introducción a la ciencia cognitiva. Barcelona: Paidós, 1990.

KAHNEMAN, D.. SLOVIC, P.. TVERSKY, A. (Eds.). *Judgment under uncertainty:* heuristics and biases. Cambridge: Cambrigde University Press, 1983.

KANT, I. *Logik.* [S.I.: s.n.|. 1800. (Redactado por Jàsche sobre apuntes de clase de Kant).

KASSIN, S.M. Base rates and prediction: the role of sample size. *Personality and Social Psychology Bulletin.* Beverly Hills (CA). v.5, p.210-213, 1979.

LEFFORD, A. The influence of emotional subject matter on logical reasoning. *Journal of General Psychology.* Provincetown (MA), v.34. p. 127-151, 1946.

LINDZEY. G„ ARONSON, E. (Eds.). *The handbook of social psychology.* 2.ed. Reading (MA): Addison Wesley, 1969. v.3: The individual in the social context.

LIPPS, T. *Grundzuge derlogik.* [S.I.: s.n|. 1893.

LOPEZ ALONSO. A.O. Test de coherencia. *Publicación CIIPME.* Buenos Aires, v.65. 1981.

------------- . Razonamiento humano: un test para su consistencia interna. *Revista Signos Universitarios.* Buenos Aires, v.7, n.13 (Especial: Psicología), 1988.

------------- . Razonamiento humano: distintos niveles de insight estructura y organización lógica subyacente (versión ampliada). *Cuadernos Pinero,* Buenos Aires, n.16, 1990.

------------- . Razonamiento humano: distintos niveles de insight, estructura y organización lógica subyacente. *Revista de Psicología General y Aplicada.* Madrid, v. 44, n.2, p.221-230,199 la.

------------- . Hacía una teoría sistémico-cibernética de la organización cognitiva de la mente y del origen de la conciencia. *Revista Signos Universitarios.* Buenos Aires, v.10, n.19 (Especial: Miscelánea I), 1991b.

--------------. Incidencia de la codificación de las premisas en el logro de un razonamiento coherente. *Revista Interdisciplinata,* Buenos Aires, v.10 n.2, p. 137-167, 1991c.

--------------. *Relational structures of premisses as determinant of reasoning coherence and thought logical organization.* Brussels, 1992. Trabajo presentado en el 25th. International Congress of Psychology, Bélgica, julio 1992.

LOPEZ ALONSO, A.O., RICARDI, N.C. La organización lògica del razonamiento en relación con categorías de diagnòstico psiquiátrico. *Revista IRICE,* Rosario, 1993.

LORD, C.R., ROSS, L, LEPPER, M.R. Biased assimilation and attitude polarization: effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology,* Washington, D.C., v.37, p.2098-2109, 1979.

MAYER, N.E., REVLIN, R. An information processing framework for research on human reasoning. In: REVLIN, R., MAYER, R.E. (Eds.). *Human reasoning.* Washington, D.C.: Winston, 1978.

McGUIRE, W.J. Cognitive consistency and attitude change. *Journal of Abnormal and Social Psychology,* v.60, p.345-353,1960.

---------------.A syllogistic analysis of cognitive relationships. In: ROSENBERG, M.J. .HOVLAND, G.I., McGUIRE, W.J., ABELSON, R.P., BREHM, J.W. (Eds.). *Attitude organization and change.* New Haven: Yale University Press, 1960.

MORGAN, J.J., MORTON, J.T. The distortion of syllogistic reasoning produced by personal convictions. *Journal of Social Psychology,* Provincetown (MA), v. 13, p. 180-184, 1944.

ORSI, A.N. Coherencia de razonamiento y diferencias individuales: un estudio de sus características e interdependencias. *Revista Signos Universitarios,* Buenos Aires, v.7, n.13 (Especial: Psicología), 1988.

PIAGET, J. *Ensayo de lógica operatoria.* Buenos Aires: Guadalupe, 1977.

POLLARD, P. Human reasoning: some possible effects of availability. *Cognition,* Amsterdan, V.12, p.65-96, 1982.

REVLIN, R. Two models of syllogistic reasoning: feature selection and conversion. *Journal of Verbal Behavior,w.l4,* p.180-195, 1975.

RICARDI, N.C. Estudio evolutivo del pensamiento y su covariación con la consistencia interna del razonamiento en niños sordos y oyentes. *Revista del Instituto de Investigaciones Educativas.* Buenos Aires, v.15, n.67/68, oct. 1989.

---------------. *La consistencia interna de razonamiento y su relación con distintas dimensiones de la personalidad, en sujetos con y sin antecedentes psiquiátricos.* Buenos Aires: Instituto de Investigaciones Psicológicas de la Universidad del Salvador, 1992. (Serie Documenta Laboris).

---------------. Construcción de premisas y coherencia de razonamiento. *IRICE-URN-CONICET,* Rosario, n.6, mar. 1993a.

----------------. Niveles diferenciales de rendimiento lógico en niños oyentes y sordos. *Interdisciplinaria,* Buenos Aires, v.ll, n.2. 1993b.

RICHTER, M. The theoretical interpretation of errors in syllogistic reasoning. *Journal of Psychology,* Provincetown (MA), v.43, p.341-344, 1957.

RIMOLDI, H.J.A. Problems of thinking about thinking: language, and logical structure in problem processing. In: FROELICH, W., SMITH, G., DRAGUNS, J., HENTSCHEL, U. (Eds.). *Psychological processes in cognition and personality.* Washington, D.C.: Hemisphere, 1984.

. A technique for the study of problem solving. *Educational and Psychological Measure ment,* Durham (NC), v.15, n.4, p.450-461, 1955.

ROMERO, F., PUCCIARELLI, E. *Lógica.* Buenos Aires: Espasa-Calpe, 1947.

SCHRODER. *Vorlesungen uber die algebra dér logik.* [S.I: s.n.], 1890.

SECHREST, L. Metodo de investigación, fiabilidad y validez. In: BELLACK, AS., HERSEN, M. (Eds). *Métodos de investigación en psicología clínica.* Bilbao: Desclée de Brouwer, 1989.

SELLS, SB. The atmosphere effect: an experimental study of reasoning. *Archives of Psychology,* V.29, p.3-72, 1936.

SMEDSLUND, J. Circular relation between understanding and logic. *Scandinavian Journal of Psychology,* Stockholm, v.ll, p.217-219, 1970.

STEBBING, L.S. *Introducción a la lógica moderna.* México: Fondo de Cultura Econòmica, 1965.

THURSTONE, L.L. Primary mental abilities. *Psychometric Monograph,* Chicago (IL): n.l, 1938.

TVERSKY. A.. KAHNEMAN, D. Judgment under uncertainty: heuristics and biases. *Science,* Washington, D.C., v.185, n.4157,p.U24-1131, 1974.

VAIHINGER. H. *La filosofia del corno si.* [S.I.: s.n.], 1878-1879.

VON DOMARUS, E. The specific laws of logic in schizophrenia. In: KASAMIN, J.S. (Ed.). *Language and thought in schizophrenia.* Berkeley: University of California Press, 1944.

VON WRIGHT, G.H. *Ensayo de lógica modal.* Buenos Aires: S. Rueda, 1970.

WASON, P.C. Problem solving and reasoning. *British Medical Bulletin. Cognitive Psychology,* Edinburgh, v.3, p.206-210,1971.

---------------. Realismo y racionalidad en la tarea de selección. In: CARRETERO, M., GARCIA MADRUGA, J.A. (Comps.). *Lecturas en psicología del pensamiento.* [S.l.]: Alianza Psicología, 1984.

----------------. Reasoning. In: FOSS. B. (Ed.). *New horizons in psychology.* London: Penguin, 1966.

---------------. Self contradictions. In: JOHNSON-LAIRD, P.N., WASON, P.C. (Eds.). *Thinking:* readings in cognitive science. Cambrigde: Cambridge University Press, 1977.

WHITEHEAD, A.N., RUSSELL, B. *Principia mathematica.* Cambrigde: Cambridge University Press, 1910.

WOODWORTH, R.S., SELLS, S.B. An atmosphere effect in normal syllogistic reasoning. *Journal of Experimental Psychology,* v.18, p.451-460, 1935.

WUNDT, W. *Ethik.* [S.I.: s.n], 1886.

---------------. *System dér philosophie.* [S.I.: s.n.], 1889.

WYER JR., R.S. *Cognitive organization and change:* an information processing approach. Potomac: Erlbaum, 1974.

CAPÍTULO 10

OBSERVAÇÃO DO COMPORTAMENTO

Cecília Guarnieri Batista
Centro de Estudos e Pesquisas Prof. Dr. Gabriel Porto (CEPRE)
Faculdade de Ciências Médicas da UNICAMP

O presente capítulo aborda os fundamentos teóricos e as diretrizes metodológicas para a realização de estudos de observação do comportamento. Exclui a pesquisa baseada em entrevistas e na análise de relatos verbais e de falas; exclui também a pesquisa relacionada à aplicação de testes e de escalas de desenvolvimento.

Observação é definida por Reber (1985), no *The Penguin Dictionary of Psychology,* com três acepções: 1) Em um sentido genérico, qualquer forma de exame de eventos, comportamentos, fenômenos, etc. 2) Por extensão, qualquer dado individual, escore, valor, etc, que represente um evento, comportamento ou fenômeno. O autor comenta que o termo "observação" muitas vezes é usado em oposição a "experimento", favorecendo a distinção entre a pesquisa com manipulação de variáveis independentes e a pesquisa que utiliza observação naturalística. Lembra, entretanto, o argumento de muitos de que tal diferenciação é desnecessária, uma vez que um experimento é meramente uma forma de fazer observação. 3) Um comentário casual ou informal a respeito de, ou uma interpretação do que foi observado.

Observação naturalística é definida, no mesmo dicionário, como a coleta de dados através de cuidadosa observação de eventos em seu ambiente natural. E **métodos observacionais** são definidos como quaisquer dos procedimentos e técnicas utilizados em pesquisa não-experimental para favorecer a observação precisa de eventos.

Este trabalho se concentra nos "métodos observacionais", utilizados especialmente em situação de "observação naturalística" do comportamento humano, de acordo com as definições acima.

ESTUDOS OBSERVACIONAIS NO BRASIL — ANOS 70

Um dos impulsos para a realização de pesquisa em Psicologia no Brasil foi a criação dos cursos de pós-graduacão na área, destacando-se

entre os primeiros os do Instituto de Psicologia da Universidade de Sao Paulo, que se iniciaram, em 1970, com a autorização dos cursos de mestrado nas áreas de concentração em Psicologia Experimental e Psicologia Escolar.

Muitos dos primeiros estudos em observação do comportamento constituíram-se nas teses de doutorado de alguns dos professores orientadores desses cursos, notadamente da área de Psicologia Experimental. Vários desses trabalhos consistiram em estudos etológicos do comportamento animal (Cunha, 1967, Carvalho, 1972 e Ribeiro, 1972, estudando formiga saúva — *Atta sexdens rubropiloser,* Forel e Ades, 1972, estudando a aranha *Arigiope argentata).* Outras teses foram desenvolvidas na área da análise do comportamento (Witter, 1969, sobre discriminação em pré-escolares; Kerbauy, 1972, sobre autocontrôlé do comportamento alimentar; Mejias, 1973, sobre modificação do comportamento escolar). Foram também realizados estudos sobre a interação mãe-criança, por docentes da USP-Ribeirão (Rossetti-Ferreira, 1967, em Londres; Marturano, 1972, na USP; Alves, 1973, na então Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto; Stella, 1974, em Londres). Destaque-se, ainda, o estudo sobre interação mãe-bebê, de Sollitto (1972) na PUC-SP. Os alunos de pós-graduacão receberam a influência desses docentes e, também, de outros professores ligados à análise do comportamento (Carolina M. Bori, Maria Amélia Matos, Margarida Windholz e Thereza P. Lemos Mettel, esta última inicialmente atuando na USP-Ribeirão e posteriormente na UnB). Dessa forma, as abordagens teóricas que mais influenciaram a realização de estudos observacionais no período foram a etologia e a análise do comportamento, que serão examinadas a seguir.

Influências teóricas e conceituais

Etologia

A etologia pode ser caracterizada como uma perspectiva biológica de estudo do comportamento. Originária da Zoologia, inicialmente caracterizou-se como uma ciência do comportamento animal. Isso levou o etòlogo, de acordo com as considerações de Carvalho (1978), a olhar o comportamento como um aspecto da Biologia do animal e a formular a respeito do comportamento o mesmo tipo de questões que formularia a respeito dos órgãos do corpo. Isso significa que o etòlogo considera que o comportamento é fruto da evolução, foi criado pela seleção natural, pela pressão seletiva do ambiente onde a espécie evoluiu e ao qual é adaptada. A autora considera que esse ponto de vista traz duas conseqüências metodológicas principais.

A primeira conseqüência se refere às questões a serem formuladas sobre o comportamento (ou sobre qualquer evento biológico) a partir dessa perspectiva, ou seja, as questões sobre o porquê causai, ontoge-

nético, funcional e evolucionário (ou filogenètico). Segundo Carvalho (1978), a pergunta causai se refere aos determinantes da ocorrência de um evento comportamental no indivíduo. A questão ontogenética se reporta à origem do comportamento no indivíduo; já a questão funcional liga-se à função que o comportamento tem para a espécie, considerando-se a análise em relação ao ambiente evolucionário do animal. Por fim, o porquê filogenètico refere-se ao modo pelo qual a evolução criou, na espécie, os mecanismos responsáveis pela emissão do comportamento no indivíduo.

Os métodos empregados para responder a essas questões, de acordo com Cunha, que se baseia em Lorenz, são, basicamente, a observação e a comparação. Para Cunha (1983), a observação é o único método realmente indispensável, mas o emprego da comparação é que tem permitido a inter-relação das quatro áreas características de investigação, referentes às questões mencionadas. Ele fundamenta essa afirmação com o seguinte argumento: de um lado, o estudo da causação do comportamento deve geralmente começar pelo levantamento do etograma de uma espécie, ou seja, "o conhecimento pormenorizado de seus comportamentos típicos em pelo menos certas fases do ciclo biológico, traduzido numa categorização e descrição com um caráter morfológico" e, de outro lado, o estudo da filogênese do comportamento geralmente envolve o estudo comparativo de duas ou mais espécies que, por outros critérios, como os morfológicos, se apresentam aparentadas. Quando se inicia o estudo filogenètico, freqüentemente se verifica a necessidade dos conhecimentos providos pelo etograma, que torna o investigador mais preparado para reconhecer características similares ou diferentes entre espécies. E essa relação de semelhança ou diferença leva à busca de uma explicação em termos de seleção natural ou de fatores de condicionamento ou treino ocorridos ao longo da vida do animal. Assim, Cunha apresenta uma visão de investigação das quatro questões, que se baseia no método de observação e que é integrada pelo método de comparação.

Voltando às colocações de Carvalho sobre conseqüências metodológicas da perspectiva biológica que caracteriza a etologia, essa autora considera que a segunda conseqüência metodológica é a ênfase no organismo. Para Carvalho (1978), cada espécie tem sua própria história evolucionaría, e cada aspecto de seu comportamento ou estrutura se torna compreensível à luz dessa história. Isso resulta em tornar extremamente importante considerar qual é o organismo que está em jogo ao se analisar o comportamento.

A etologia dedicou-se, inicialmente, ao estudo do comportamento de diferentes animais (insetos, peixes, aves, mamíferos, notadamente primatas). Os estudos do comportamento humano passaram a aumentar a partir dos anos 70 (Hutt e Hutt, 1974, original de 1970; Blurton Jones, 1972;

McGrew, 1972). Estes estudos têm ênfase descritiva, trazem em geral relações extensas de categorias de comportamentos definidos morfologicamente, utilizam como medidas a freqüência de ocorrência, a duração e a intensidade, e buscam efetuar análise de seqüência de categorias.

Em nosso meio, Cunha (1974) ministrou por vários anos um curso de observação do comportamento animal de orientação etològica, formalizando uma série de aspectos referentes à elaboração de definições de comportamento. **Comportamento** é entendido como um fenômeno biológico, "constituído de **estados orgânicos, posturas** e **movimentos** cuja forma e organização não podem ser atribuídas diretamente a energias externas ao sistema orgânico, que os exibiria como uma manifestação de sua irrita— bilidade e modificabilidade" (Cunha, 1975, p.264). Os **estados orgânicos** são "revelados por aspectos funcionais tais como temperatura, cor, rigidez, etc, de uma estrutura orgânica ou suas partes"; as **posturas** são definidas como "disposições espaciais estacionarias de partes do organismo umas em relação a outras"; e os **movimentos,** como "mudanças de posição espacial de uma ou mais partes do organismo em relação a outras, tomadas como referência" (id. ibid., p.265).

Análise do comportamento aplicada

A análise do comportamento aplicada teve suas bases definidas ao final dos anos 60 e foi derivada da análise experimental do comportamento, que aborda a relação comportamento-ambiente corno em constante alteração recíproca, conforme se depreende da afirmação de Skinner (1975, p.1) de que "os homens agem sobre o mundo e o transformam e por sua vez são transformados pelas conseqüências de suas ações". As unidades de estudo são constituídas por classes de estímulo, definidas através de seus efeitos no comportamento, e por classes de respostas, definidas por suas conseqüências no ambiente (Catania, 1973). As relações entre esses eventos são explicitadas na contingência de reforçamento: segundo Skinner (1975, p. 12), "uma formulação das interações entre um organismo e o seu meio ambiente, para ser adequada, deve sempre especificar três coisas: 1) a ocasião na qual ocorreu a resposta, 2) a própria resposta, e 3) as conseqüências reforçadoras. As relações entre elas constituem as contingências de reforçamento". A estratégia metodológica da análise experimental do comportamento é explicitada por Sidman (1960) como estando baseada no estudo de alterações na freqüência acumulada de uma resposta ao longo do tempo, identificando-se a influência da introdução de variáveis sobre uma linha de base estável. Dessa forma, a análise experimental do comportamento enfatiza a importância do estudo das mudanças no comportamento à medida em que ocorrem, em contraposição a outras abordagens experimentais que se limitam à medição anterior e posterior à realização da intervenção experimental.

A análise do comportamento aplicada foi delimitada por Baer, Wolf e Risley (1968, p.91) corno "o processo de aplicar princípios de comportamento, algumas vêzes em experiência, para a melhoria de comportamentos específicos, e de avaliar simultaneamente se as mudanças observadas poderiam ser atribuídas ao processo de aplicação — e, em caso positivo, a que partes do processo". Assim, a análise aplicada está dirigida para o exame de comportamentos que são socialmente relevantes, ao invés de serem convenientes para estudo. Segundo esses autores, o estudo deve ser **aplicado, comportamental** e **analítico;** além disso, deve ser **tecnológico, conceitualmente sistemático** e **efetivo,** devendo, ainda, demonstrar alguma generalidade.

Quanto à metodologia de coleta de dados, Bijou, Peterson e Ault (1968, p. 175) propõem "um método para integrar os campos de estudo descritivo e experimental a nível dos dados e dos conceitos empíricos". Eles consideram que essa integração é possível se for utilizada como dado básico a freqüência de ocorrência de eventos ambientais e comportamentais, sendo que a informação sobre as relações funcionais entre eventos só poderão ser fornecidas por estudos experimentais. Os autores estabelecem procedimentos para a condução de estudos descritivos de campo, incluindo as seguintes considerações: a) **Especificação da situação em que o estudo é conduzido** — segundo os autores, a situação deve ser definida em termos do ambiente físico e social e dos eventos observáveis que ocorrem no mesmo, buscando-se o mínimo de variações na mesma ao longo do estudo, b) **Definição de eventos comportamentais e ambientais em termos observáveis** — os autores consideram que a seleção de eventos comportamentais e ambientais deve ser derivada de estudos preliminares, a partir dos quais se constrói um código observacional, consistindo de itens definidos e de seus símbolos. Os eventos comportamentais deverão ser registrados em sua freqüência real ou em termos de sua ocorrência ou não ocorrência a intervalos de tempo prefixados. Eventos ambientais podem se referir a objetos, descritos em termos de suas características físicas; a eventos biológicos, difíceis de medir em situações de campo; e a eventos sociais, definíveis do mesmo modo que definem eventos comportamentais, e) **Mensuração da fidedignidade do observador** — os autores apresentam a fórmula para cálculo do acordo entre observadores (acordos/acordos + desacordos) e propõem procedimentos para evitar a ocorrência de desacordos, d) **Procedimentos para coleta, análise e interpretação de dados** — os autores sugerem que a coleta de dados seja realizada a intervalos regulares, que a análise consista na transformação dos dados de modo a facilitar a visualização das relações entre eventos e que a interpretação se limite a conceitos empíricos e relações consistentes com as observações e análises efetuadas. Eles consideram que, em um estudo descritivo de campo, as interpretações geralmente consistem em uma comparação do

que foi descoberto na situação com outros dados obtidos sob condições funcionalmente semelhantes.

Diretrizes metodológicas

Como decorrência das proposições teóricas da etologia e da análise do comportamento aplicada, foi desenvolvida uma série de diretrizes metodológicas para a realização de estudos observacionais, que serão descritas a seguir. Antes, porém, é importante lembrar a questão da seletividade. Segundo Hinde (1966), a seleção na observação é inevitável: a descrição e a classificação envolvem rejeição de dados e seleção de critérios. Sempre se corre o perigo de rejeitar os dados essenciais e utilizar unidades de análise irrelevantes. No entanto, a rejeição insuficiente de dados pode levar a um conjunto de detalhes confusos e desviar a atenção dos aspectos essenciais do problema. Para Hinde (1973), o grau de seleção depende do problema estudado e do fato de que a precisão do registro diminui à medida que aumenta o número de respostas a registrar.

Delimitação de unidades e definição de categorias de comportamento

Hinde (1973) lembra que o comportamento se apresenta numa complexa sucessão de eventos no tempo: é preciso dividi-lo em unidades para possibilitar sua descrição e classificação. As unidades devem ser delimitadas a partir de pontos de transição, ou seja, de fenômenos que se distinguem, **por sua forma ou por seus resultados,** do fenômeno imediatamente precedente (Cunha, 1975). A divisão do comportamento em unidades, a partir de **mudanças em sua forma** (descrição física ou morfològica), tem sido enfatizada pelos etólogos (Blurton Jones, 1972; Cunha, 1976). Na divisão do comportamento **pelos seus resultados** (descrição pela conseqüência ou funcional), separa-se o contínuo da ação cada vez que o comportamento produz um efeito particular sobre o meio. Hinde (1966) aponta duas vantagens básicas da definição funcional: a) uma breve descrição pode cobrir um grande número de padrões motores variáveis; b) esse tipo de descrição chama a atenção para características essenciais do comportamento, que podem não aparecer em descrições pela forma, tais como a orientação com respeito ao ambiente ou a qualidade de responder a estímulos externos. Apesar de alguns etólogos ressaltarem a importância da definição pela forma, outros afirmam que ambos os tipos de definição podem ser usados e mesmo combinados num catálogo (Hinde, 1973; Hutt e Hutt, 1974). Segundo Hinde (1973), a preferência depende da natureza do comportamento em questão, bem como do objetivo da pesquisa. Danna e Matos (1982) consideram que definições morfológicas são mais adequadas quando o comportamento não produz mudanças perceptíveis no ambiente externo, como é o caso, por exemplo, de mudanças de expressão facial e de postura. Consideram, ainda, que tais definições são necessárias quando se pretende estudar o funciona-

mento de determinada parte do corpo e quando se visa ao treinamento, como no caso da habilitação de pessoas com deficiência física. Já as definições funcionais são suficientes quando a observação visa à seleção ou avaliação de pessoas e preferíveis quando os aspectos morfológicos consistirem basicamente de movimentos de difícil identificação e observação, como, por exemplo, as vocalizações.

Uma vez dividido o fluxo de eventos comportamentais em unidades, é preciso agrupá-los em categorias, de acordo com as propriedades comuns que os caracterizam, e em seguida defini-los. Apresentam-se, a seguir, algumas **regras para a definição de categorias de comportamento:**

a)**A** definição deve indicar aquelas propriedades que distinguem uma categoria comportamental de outras categorias. Deve especificar uma ou mais regras gerais que explicitem essas propriedades, demarcando os limites do que deve ser incluído naquela categoria ou excluído da mesma (Michael, 1975).

b)A definição deve ser objetiva, referindo-se a comportamentos observáveis tão explicitamente quanto possível (Michael, 1975). Não deve conter inferência, interpretação ou finalismo (Cunha, 1974).

c)A definição deve ser clara, evitando-se a linguagem ambígua (Michael, 1975). Diz-se que um termo é ambíguo quando tem dois significados distintos e o contexto não esclarece qual dos dois está sendo utilizado (Copi, 1974).

d)A definição não deve ser circular, ou seja, o termo a ser definido, ou um sinônimo dele, não pode figurar no grupo de elocuções destinadas a esclarecer o significado desse termo (Copi, 1974).

e)A definição deve apontar um referencial de comparação quando utilizar um adjetivo sem significado absoluto, tal como largo, alto, pequeno, etc.

f) A definição pode incluir alguns exemplos, especialmente de inclusões e exclusões difíceis de julgar (Michael, 1975).

g)A definição deve especificar o critério para delimitação de unidades, de modo a permitir a contagem.

Uma vez definida a categoria comportamental, deve-se atribuir-lhe um nome que, para Cunha (1976), deve evocar a definição da categoria em termos de observáveis, reduzindo a possibilidade de se atribuir à categoria outros significados que não aqueles que determinaram sua criação. O autor considera ser interessante atribuir um nome a uma categoria já definida, pois desse modo podemos nos referir a ela sem precisar repetir sua definição.

Elaboração do catálogo de categorias de comportamento

O termo "categoria de comportamento", como foi visto anteriormente, designa o conjunto de unidades comportamentais agrupadas de acordo com propriedades comuns. E o termo "catálogo" será empregado para designar o conjunto de categorias empregadas num determinado estudo. Em um catálogo se procura descrever, de modo completo e sistemático, o repertório comportamental dos sujeitos observados, no que se refere ao objetivo do estudo em questão. Pode também ser incluída no catálogo a relação de circunstâncias em que as categorias são observadas.

Discutem-se, a seguir, alguns problemas e princípios de categorização e de elaboração de um catálogo.

Uma categoria do catálogo não deve ser excessivamente ampla, incluindo sob a mesma denominação um conjunto de eventos cuja unicidade ainda não se demonstrou. Uma crítica à inadequação metodológica do uso de tais categorias ("variáveis resumidoras"), no estudo da interação mãe-criança, foi feita por Marturano (1972). Ao se agruparem diferentes eventos sob a mesma categoria, deve-se perguntar se se trata de uma espécie de comportamento ou de mais de uma espécie. Em relação a esse problema, Altmann (1965) sugere separá-los quando em dúvida, pois mais tarde sempre se poderá juntar duas categorias em uma só, mas não se poderá separar o que se juntou precocemente.

Certos cuidados durante a categorização justificam-se por suas implicações para posterior análise de dados. Slater (1973), ao discutir análise de dados, apresenta como exigências: a) que as categorias sejam mutuamente exclusivas, e b) que o observador esteja preparado para tratar todos os membros de uma dada categoria como equivalentes.

Um catálogo deve, também, permitir diferentes níveis de análise do fenômeno, do mais molecular ao mais molar. Uma forma de atender a essa exigência é elaborá-lo em termos de categorias mais amplas e de subcategorias que as compõem. Os dados são coletados com base nas subcategorias, mas podem também ser agrupados nas categorias mais amplas, de modo a se ter uma visão de conjunto do fenômeno.

A habilidade na elaboração de categorias depende muito da experiência prévia do pesquisador, tanto no que se refere ao treino anterior em observação como na sua experiência específica com os comportamentos sob análise. Uma vez elaboradas as categorias, é interessante que ele as discuta com outros pesquisadores experientes. Dois critérios objetivos foram propostos por Richards e Bernal (1972), para testar a adequação das categorías: a) Fidedignidade e validade — diferentes pesquisadores tendem a desenvolver categorias idênticas ou semelhantes, e uma descrição da categoria permitirá que outro observador a registre e obtenha um alto grau de concordância; b) Utilidade na análise — a ocorrência da

categoria deve estar correlacionada com outros eventos ou mostrar distinções entre grupos pré-selecionados de sujeitos.

Muitas vezes, ao se elaborar um catálogo, agrupa-se uma série de eventos considerados pouco relevantes sob uma categoria residual ou indiferenciada. Weick (1968) lembra que é preciso evitar, entretanto, que um número significativo de eventos incida nessa categoria, pois nesse caso o sistema de classificação se tornaria difícil de interpretar.

Um exemplo de catálogo de comportamentos fundamentado no critério morfológico é o de Vieira (1975). A autora observou pessoas em diferentes situações e elaborou um catálogo com 146 itens, incluindo unidades anatômicas (por exemplo, "íris", "pàlpebra") e de comportamento (por exemplo, "abrir os olhos", "punho cerrado"), tomando como referencial o próprio corpo e sua posição no espaço, num trabalho diretamente derivado das proposições de Cunha (1974, 1975, 1976).

Um outro exemplo é o catálogo elaborado por Batista (1978, 1980), a partir da observação de pais e filhos pré-escolares em situação de refeição. O catálogo consta de 123 verbos definidos e de uma relação de termos da oração que devem ser registrados quando da utilização de cada um desses verbos (por exemplo: quando for empregado o verbo "abaixar", indicar sujeito, objeto direto e adjunto adverbial que descreve até onde o objeto foi abaixado). Cada definição está redigida sob a forma de uma contingência de três termos — condição, resposta, produto. As relações entre verbos definidos e definidores está representada em diagramas de árvore, o que favorece a autocorreção de cada definição e do catálogo como um todo, dificultando a ocorrência de categorias que não sejam mutuamente exclusivas. Esse estudo representa um exemplo da influência dupla de etologia e da análise do comportamento: elabora-se um extenso catálogo de comportamentos, como nos estudos etológicos, ao mesmo tempo que se formulam definições predominantemente funcionais. Analisando-se o trabalho em retrospectiva, verifica-se que o catálogo teve generalidade, tendo sido utilizado por outros autores em outras situações. Entretanto, considera-se que a abordagem à situação que se pretendia estudar poderia ter sido feita com a definição prévia de objetivos mais específicos e menos exploratórios. Considera-se, agora, de duvidosa utilidade a elaboração de catálogos genéricos de comportamentos, sem que estejam pautados pelo mesmo enfoque teórico que leva à elaboração de etogramas.

Descrição das circunstâncias em que os comportamentos ocorrem

Ao apresentar princípios de categorização e análise do comportamento dentro de uma perspectiva etològica, Cunha (1976) fala sobre a necessidade de especificação e descrição das circunstâncias em que ocorre o comportamento, e afirma que elas podem ser encontradas entre

os aspectos do ambiente, do organismo e de seu comportamento. Para ele, entre os aspectos do **ambiente,** interessam especialmente as mudanças no ambiente imediato do organismo que se correlacionam com seu comportamento. Entre os aspectos do **organismo,** incluem-se sua espécie, o estágio de desenvolvimento em que se encontra, a condição presente de seus vários órgãos, a postura do animal como um todo ou de alguma de suas partes, mudanças que podem estar ocorrendo no interior do organismo em virtude de ter sido exposto a determinadas circunstâncias ambientais no passado. E entre os aspectos do **comportamento,** podem ser mencionados os comportamentos concomitantes ao comportamento observado, os comportamentos que o precedem imediatamente e os comportamentos passados do animal diante de situações similares ou mesmo diferentes da observada (Cunha, 1976).

No estudo observacional realizado em situação natural só é possível estudar diretamente as circunstâncias que se referem a ambiente, organismo e comportamento, na medida em que puderem ser observadas sem o auxílio de instrumentos ou de contato físico com o organismo estudado. Através de entrevista podem-se obter alguns outros dados, tais como: idade; condição física passada e presente; condição de privação alimentar, de atividade e de sono; história de influência ambiental, como, por exemplo, separação prematura dos pais, o fato de ter sido submetido a tratamentos referentes à saúde física e mental, escolaridade, nível sócio-econômico, etc.

A questão da identificação das circunstâncias do comportamento também é relevante para a análise do comportamento. Ao propor uma análise funcional do comportamento em 1953, Skinner (1967) enfatiza as condições externas das quais o comportamento é função e os experimentos realizados segundo o paradigma operante envolvem a manipulação de estímulos apresentados em estreita contigüidade temporal com a resposta. Por outro lado, Staddon (1973) afirma que não se justifica a tentativa de explicar o comportamento, limitando-se a apontar como causas apenas estímulos temporalmente contíguos. O autor considera que um comportamento poderá ter muitas causas, e que todas ou quase todas poderão estar no passado. É possível que estejam ligadas umas às outras e ao resultado final através de mecanismos de imensa complexidade. O que falta é superar as dificuldades no estudo dessas influências.

Um estudo observacional terá limitações na identificação de todos os eventos ambientais que possivelmente estejam influindo no comportamento. É importante lembrar que a identificação de eventos temporalmente contíguos ao comportamento e a descrição do contexto ambiental, sugerindo testes experimentais, se constituem soluções parciais para a questão.

Escolha da técnica de registro

Os autores ligados à análise do comportamento aplicada definiram uma série de técnicas de registro que permitissem a coleta de dados relativas a classes de respostas predefinidas, sua quantificação e representação gráfica. Hall, um autor clássico na área, distingue as seguintes técnicas de mensuração: **registro automático, mensuração direta de produtos permanentes** e **registro de observação. O registro automático** envolve o uso de instrumentos nos quais "a resposta do organismo aciona um dispositivo elétrico ou mecânico que, por sua vez, registra automaticamente cada resposta" (Hall, 1973, p.2). O autor destaca como vantagens de seu uso a precisão e a facilidade de tradução do resultado em termos numéricos e como desvantagens aspectos ligados a custo e manutenção, bem como à falta de flexibilidade ou de capacidade de registrar muitos comportamentos específicos de interesse do pesquisador. A **mensuração direta de produtos permanentes** envolve a análise de resultados do comportamento (por exemplo, respostas escritas a tarefas acadêmicas, blocos empilhados, etc), sendo destacadas como vantagem a precisão e a possibilidade de quantificação. Já o **registro de observação** ocorre quando "um observador humano vê um comportamento e o registra na forma como ocorre" (id. ibid., p.3). O autor destaca os seguintes tipos de registro: contínuo, de evento, de duração, de intervalo, amostragem de tempo e *placheck,* descritos a seguir:

— registro contínuo: caracterizado como aquele em que "o observador tenta escrever tudo o que está ocorrendo" (id. ibid, p.3), com a vantagem de possibilitar a inclusão de muitas classes de comportamento e a desvantagem de tomar muito tempo do observador, segundo o autor.

— registro de evento: envolve a "contagem da freqüência de eventos na medida em que estes ocorrem" (id. ibid., p.3-4), através de marcas no papel ou contadores de pulso, sendo indicado como vantagem o fato de não interferir muito nas tarefas em andamento.

— registro de duração: envolve o registro da "duração de determinado comportamento durante um dado período de observação" (id. ibid., p.4), geralmente através de cronômetro.

— registro de intervalo: para esse tipo de registro, "cada sessão de observação é dividida em períodos de tempo iguais. O observador registra então a ocorrência do comportamento durante esses intervalos" (id. ibid., p.4). Assim, para cada intervalo, é feita uma marca correspondente à ocorrência do comportamento, tenha sido ele de curta ou longa duração ou tenha ocorrido uma ou mais vezes. Essas questões em geral são resolvidas delimitando-se intervalos bastante pequenos, em torno de 10 a 15 segundos. O autor aponta como vantagens o fato de a técnica

dar indicação tanto da freqüência quanto da duração do comportamento observado, podendo-se registrar simultaneamente mais de um tipo de comportamento. Aponta como desvantagem a exigência da atenção ininterrupta do observador.

— *placheck:* de *Planned Activity Check,* envolve a contagem, a intervalos predeterminados, do total de indivíduos que estão desempenhando o comportamento alvo e do total de indivíduos presentes, exprimindo-se o resultado em porcentagem.

Hall (1973) considera que os registros de observação em geral não são tão precisos quanto os obtidos com instrumentos automáticos, mas que são vantajosos para utilização em situações de aplicação de modificação de comportamento, como a escola e o lar.

Uma revisão mais recente de técnicas de registro foi realizada por Fagundes (1981), em que o autor dá exemplos detalhados da aplicação de cada técnica e introduz algumas técnicas mistas, além das tradicionalmente utilizadas.

Embora autores norte-americanos da área de análise do comportamento aplicada considerem que a técnica de registro contínuo só é útil nas fases preliminares do estudo, esta tem sido bastante utilizada no Brasil como técnica básica de coleta de dados. Os primeiros estudos da interação mãe-criança são um exemplo disso. Sollitto (1972) estudou a interação mãe-nenê durante o banho: as respostas verbais da mãe e do nenê foram registradas pelo gravador, e as respostas motoras foram anotadas por escrito, de modo contínuo, separadas por intervalos de 10 segundos. A gravação foi transcrita e elaborou-se um protocolo de cada sessão de observação, unindo-se o registro das respostas motoras da mãe e do nenê às transcrições das respostas verbais de ambos (do bebê, através de transcrição fonética). A análise de dados envolveu o cómputo de freqüências de respostas e a elaboração de gráficos de freqüência acumulada das mesmas. Marturano (1972) estudou a interação mãe-criança de três anos em três condições que envolviam a realização de uma tarefa — refeição em casa, refeição no laboratório e resolução de quebra-cabeças em laboratório — e utilizou técnica de registro e transcrição de dados semelhante à de Sollitto (1972). A análise de dados consistiu no computo de freqüências e na análise de seqüência, esta envolvendo a tabulação das unidades verbais em matrizes de seqüências, análise probabilistica das seqüências observadas e análise probabilistica dos padrões de seqüência.

Esses estudos ilustram o emprego da técnica de registro cursivo como base para a realização de análise quantitativa. Tem sido apontado como vantagem da utilização dessa técnica o fato de se evitar a seleção prévia de variáveis, como seria o caso se fosse adotado um catálogo de categorias, bem como de preservar a seqüência dos eventos observados.

Por outro lado, há alguns cuidados a serem tomados. Altmann (1974), ao discutir diferentes técnicas de amostragem para estudo do comportamento social, com enfoque predominante em colônias de primatas, dá o nome de "amostragem *ad libitum"* à situação em que o observador "registra tanto quanto consegue" ou aquilo que é mais prontamente observável no comportamento social de um grupo. Ela aponta algumas dificuldades dessa forma de trabalho — segundo ela, diferentes comportamentos de diferentes sujeitos não têm a mesma probabilidade de serem registrados. Além disso, mesmo que a probabilidade de registro se mantivesse a mesma, em muitas ocasiões o observador veria mais eventos do que conseguiria registrar e, na ausência de critérios sistemáticos de seleção, suas preferências pessoais governariam a escolha. A "amostragem *ad libitum"* não é totalmente equivalente à técnica de registro cursivo, e, nos estudos de interação acima citados, houve um cuidado das autoras na padronização das condições de registro e da linguagem a ser utilizada, antes da realização da coleta de dados. Fica, entretanto, a recomendação: há necessidade de algumas definições sobre sujeitos e situações de registro, bem como de padronização da linguagem, antes da realização de uma coleta de dados empregando registro cursivo, principalmente se vai se efetuar uma análise quantitativa dos dados.

Ainda sobre a técnica de registro cursivo, urna série de orientações sobre como realizá-lo são apresentadas em Danna e Matos (1982). E, no que se refere à elaboração de listas de assinalar *(check-sheets),* Hinde (1973) apresenta uma série de sugestões e discute as vantagens e desvantagens de sua adoção.

Análise quantitativa de dados

A análise quantitativa mais elementar envolve contagens de freqüência, geralmente convertidas em **freqüência relativa** ou porcentagem. Autores ligados à análise do comportamento têm proposto preferencialmente a representação dos dados em gráficos, que permitem visualizar alterações de freqüência ao longo do tempo e envolvem um mínimo de manipulação estatística dos dados.

Com relação à etologia, Slater (1973) afirma que a **análise de seqüência** tem sido muito enfatizada por etólogos, especialmente os interessados na causação do comportamento. Em geral, eles consideram que a ocorrência de dois padrões de comportamento em proximidade temporal indica que eles compartilham alguns fatores causais, ou porque ambos os comportamentos dependem de um estado corporal particular (a presença de um hormônio, por exemplo), ou porque eles podem ocorrer como resposta ao mesmo estímulo externo ou a estímulos externos relacionados. O autor destaca que a análise de seqüência não pode fazer a distinção entre essas possibilidades, mas pode indicar de forma objetiva

os agrupamentos em que os comportamentos ocorrem e assim definir as relações que precisam ser explicadas. O autor distingue **métodos de análise de seqüências de comportamento intraindividual** do **estudo de seqüências de comportamento em situação de interação social.** Quanto aos métodos de **análise de seqüência intraindividual,** Slater (1973) distingue dois tipos: a) análise de freqüências de transição de probabilidades condicionais e b) comparação com um modelo randômico. Quanto ao primeiro, o autor afirma que o tipo mais simples de seqüência de eventos é a seqüência determinista, embora a maioria das seqüências de comportamento sejam probabilísticas e não deterministas. Seqüências probabilísticas bastante precisas são usualmente designadas como "respostas encadeadas". Nesses casos, a probabilidade de um evento particular é tão marcantemente alterada pela natureza do evento imediatamente precedente que um diagrama de fluxo indicando as freqüências com que ocorrem as diferentes transições prove uma boa impressão da organização do comportamento (por exemplo, as seqüências de cortejo em vertebrados inferiores). Quando as seqüências não são tão ordenadas, pode-se incluir apenas as transições que têm alta probabilidade de ocorrência (probabilidade condicional da ocorrência de B, dado que A acabou de ocorrer). O autor afirma que esse tipo de análise só **é** útil quando a seqüência é forte e os diferentes comportamentos considerados ocorrem com freqüências semelhantes.

O segundo método de análise de seqüência intraindividual, comparação com um modelo randômico, envolve as cadeias de Markov. "Uma seqüência de comportamento pode ser descrita como uma cadeia de Markov se as probabilidades de diferentes atos dependerem apenas do ato imediatamente precedente e não de nenhum ato anterior" (Slater, 1973, p.135). Tal seqüência será referida como uma cadeia de Markov de I[8] ordem, sendo que "uma cadeia de Markov de rª ordem é aquela em que um evento específico é afetado significativamente pelos **r** eventos precedentes" (id. ibid., p. 135). Grande parte das análises de seqüência do comportamento envolvem a identificação de dependências de 1ª ordem, a partir da comparação da matriz de freqüências de transição observadas com as freqüências esperadas, caso os atos fossem independentes um do outro. O autor apresenta uma série de estratégias e cuidados a serem tomados para a aplicação desse tipo de análise. Como métodos alternativos, menciona a correlação e a análise fatorial e alerta para o fato de que não se sabe claramente o que representam as variáveis hipotéticas extraídas por esse último método.

Todas as considerações feitas até o momento por Slater (1973) referem-se às seqüências de comportamento intraindividual que ocorrem em contextos em que se supõe que o mundo externo permaneça relativamente imutável. Isso é diferente do estudo de seqüências de comportamento em situação de interação social, que tem um objetivo primário

diferente: demonstrar que o comportamento de um animal é afetado pelo de outros e obter evidência do papel de diferentes padrões de comportamento na comunicação. Uma dificuldade apontada para essa análise é o fato de que os dados podem não ser estacionarios, uma vez que muitos comportamentos interessantes na comunicação são aqueles cuja probabilidade muda à medida que a interação prossegue. Uma outra dificuldade é que o comportamento de um animal em situação social depende, em parte, da seqüência de atos que ele já apresentou, bem como do comportamento de outros. O modelo de análise de interação entre A e B deveria levar em conta, segundo o autor, seqüência de comportamentos de A, seqüência de comportamentos de B, seqüência de comportamentos de A dirigidos a B e de B dirigidos a A, levando-se em conta que os eventos imediatamente antecedentes não são os únicos a afetar um determinado comportamento (Slater, 1973).

Verifica-se, assim, que o estudo de seqüências de comportamento em situação de interação social demanda uma atenção especial, seja em termos da própria abordagem ao fenômeno, seja em termos da análise quantitativa do mesmo. Alguns progressos recentes em relação à análise quantitativa de dados são apresentados por Van Hoof (1982), Sato (1987) e Lopes, Bueno e Barnabe (1992).

Avaliação da fidedignidade

O conceito de fidedignidade, de acordo com a teoria clássica dos testes, é apresentado por Johnson e Bolstad (1973, p.26) como "a consistência com que um teste mede um dado atributo ou fornece um escore consistente em uma dada dimensão", sendo que "a exigência clássica de fidedignidade envolve a consistência no instrumento de medida ao longo do tempo (fidedignidade no teste-reteste) ou ao longo de conjuntos de itens respondidos na mesma ocasião (fidedignidade *split-half)"* (id. ibid., p.10). O autor discute a aplicação do conceito aos estudos observacionais em análise do comportamento aplicada e argumenta que o cálculo do acordo entre observadores não deve se constituir a única forma de aferir a fidedignidade desses estudos.

Weick (1968) cita Medley e Mitzel, que sugerem três tipos de fidedignidade, calculados através de diferentes índices: acordo entre observadores (diferentes observadores observando ao mesmo tempo), estabilidade (o mesmo observador observando em ocasiões diferentes) e fidedignidade (observadores diferentes observando em ocasiões diferentes). A medida de acordo entre observadores, mais comumente utilizada, privilegia a intersubjetividade, enquanto as demais privilegiam a replicabili dade. O autor também cita Dunnette, que argumenta que há várias fontes de erros quando se faz observação, e que o tipo de fidedignidade necessária para avaliar ou compensar cada um desses erros é diferente.

Os tipos de erros mencionados são: a) **amostragem inadequada de conteúdo,** quando observadores diferentes colhem amostras de apenas alguns elementos de um comportamento complexo e essas amostras consistem de elementos diferentes; b) **tendências de respostas ao acaso,** que provêm de definições imprecisas de categorias ou de compreensão inadequada da categoria por parte do observador, que o levam freqüentemente ao uso de definições intuitivas informais, mais variáveis que as formalmente definidas; c) **mudanças no ambiente** e **d) mudanças na pessoa** que está sendo observada, duas alterações que podem ocorrer ao longo do período de observação. Dadas essas várias fontes de erro, Weick (1968) sugere que a fidedignidade seja medida de vários modos. Num estudo ideal de observação, seriam feitas quatro comparações: 1) as classificações de duas pessoas observando o mesmo evento seriam correlacionadas, prevenindo erros de mudanças na pessoa e no ambiente; 2) as classificações de um mesmo observador observando um evento semelhante em duas ocasiões diferentes seriam comparadas, evitando-se erro de amostragem de conteúdo; 3) depois seria correlacionado o acordo de dois observadores observando um evento em duas ocasiões diferentes, medida vulnerável às quatro fontes de erro, da qual se pode esperar a fidedignidade mais baixa dentre as quatro comparações; 4) finalmente, as observações de um único observador observando um único evento seriam comparadas às outras correlações, em uma verificação da consistência interna do observador. O autor considera que é possível apenas aproximar-se deste ideal de avaliação da fidedignidade, priorizando-se o acordo entre observadores em relação a um único evento. Tendo em vista as atuais possibilidades de gravação em vídeo, considera-se que seria bastante viável adaptar essas propostas para avaliação da fidedignidade tanto durante o treino de observadores como no teste de categorias de comportamento.

Uma revisão das variáveis que podem afetar o índice de concordância foi realizada por Batista (1977). No que se refere ao acordo entre observadores quando é utilizada a técnica de registro cursivo, Batista e Matos (1984) discutem a questão e sugerem definições e medidas aplicáveis a essa técnica de registro.

Batista (1985) analisa as diferenças de abordagem à questão da fidedignidade por autores ligados à etologia e à análise do comportamento e conclui que o pesquisador deve adequar a avaliação de fidedignidade às características específicas de seu projeto de trabalho. A autora faz algumas sugestões específicas:

a) Estudos preliminares que visem à caracterização de um fenômeno podem se beneficiar mais de discussões informais entre pesquisadores do que de testes formais de fidedignidade.

b) Estudos em que um único observador efetua todos os registros de comportamento podem ter avaliações de fidedignidade de várias

formas, entre as quais 1) a aferição da estabilidade do observador através de verificação de seu desempenho no registro de um mesmo vídeo em duas situações diferentes e 2) a comparação de seu desempenho com o de um pesquisador experiente que tenha tido contato prévio com as definições das categorias apenas por escrito.

c) Estudos em que diferentes observadores participem da coleta de dados terão que contar necessariamente com um sistema de aferição da precisão de cada um desses observadores. Um alto índice de acordo, no entanto, não indicará necessariamente a adequação do sistema de categorias utilizado, uma vez que estes poderão estar reagindo a uma definição implícita da categoria.

d) O teste das definições de categorias poderá ser realizado, em parte, através da análise dos desacordos constatados para categorias específicas.

No que se refere à fidedignidade, o importante é reter o significado mais amplo do termo e estabelecer formas de aferi-la que sejam compatíveis com os objetivos de cada estudo.

ESTUDOS OBSERVACIONAIS NO BRASIL — ANOS 80 E 90

Ao longo dos anos 80 e 90, foram criados vários centros de formação de pesquisadores no País. Novas perspectivas teóricas vieram se juntar às já existentes, norteando a realização de estudos observacionais. Esses estudos têm sido apresentados e discutidos em reuniões científicas, destacando-se a Reunião Anual da SBPC — Sociedade Brasileira para o Progresso da Ciência — e a Reunião Anual da Sociedade Brasileira de Psicologia (sucessora da Sociedade de Psicologia de Ribeirão Preto), e publicados em revista como *Psicologia, Psicologia: Teoria e Pesquisa, Boletim de Psicologia* e outras. Entre as publicações figuram também dois livros didáticos sobre utilização de metodologia observacional (Fagundes, 1981; Danna, Matos, 1982), um livro com uma coletânea de estudos de observação da interação professor-aluno (Morais, 1980) e um relato de pesquisa sobre apego, combinando observação direta do comportamento em situação planejada e dados obtidos através de entrevistas e visitas domiciliares (Rossetti-Ferreira, 1986).

Em um *workshop* sobre questões metodológicas no estudo da interação humana, um grupo de autoras (Alves, Carvalho, Marturano, Mettel, Rossetti-Ferreira, 1987) apresentaram os resultados de um levantamento dos estudos utilizando metodologia observacional ao longo de quinze anos (levantamento dos resumos de pesquisas apresentadas à Sociedade de Psicologia de Ribeirão Preto entre 1971 e 1986). Verificaram que houve um aumento gradual no total de trabalhos e que esse aumento

ficou mais visível a partir de 1980. Nas primeiras reuniões houve um predomínio de estudos com animais e, ao longo do tempo, verificou-se um aumento e diversificação de estudos do comportamento humano. Atécnica de registro mais utilizada foi a de registro contínuo (50%), seguida pelo de categorias (23,9%) e de *check-list{l0,2%).* Os recursos mais utilizados para registro foram papel e lápis (4ª,9%), seguidos de gravador (23,9%) e videocassete (15,9 %).

Nesse *workshop* foram discutidas várias questões. A primeira delas referiu-se ao conceito de interação, encarada como algo que ocorre entre os sujeitos, de forma que a linguagem adequada à descrição do comportamento de indivíduos não é adequada ao estudo da interação entre eles. Quanto aos métodos de coleta de dados, enfatizou-se a necessidade de definição e explicitação dos objetivos dos projetos, pois deles estariam realmente dependendo todas as escolhas posteriores. Quanto ao uso de hipóteses, entendidas como construto, conceituação do fenômeno, considerou-se que, no Brasil, os pesquisadores nem sempre estabelecem explicitamente o nexo entre sua hipótese e o recorte que vão fazer do processo de interação. Lembrou-se, ainda, que o pesquisador deve estar alerta para o fato de que esse recorte ocorre tanto na coleta como na análise de dados. Constatou-se uma tendência na pesquisa da interação caracterizada por: a) um deslocamento do enfoque molecular na direção do enfoque molar; b) uma ênfase crescente no contexto onde a interação se processa; c) o desenvolvimento de formas de recorte, a partir do recorte antecedente-conseqüente, na direção de unidades mais amplas, como episódio. Para alguns participantes do *workshop,* uma das questões centrais referiu-se à sistematização das observações, de forma a poder transformar seu produto em teoria ou generalizar os dados obtidos (Alves, Carvalho, Marturano, Mettel, Rossetti-Ferreira, 1987, com redação de Alves e Marturano).

Preocupação semelhante com o direcionamento da coleta e análise de dados foi manifestada por Prado, criticando um certo tipo de estudo observacional, para ela inadequadamente denominado "etològico", que partiria da afirmação de que a Psicologia é uma ciência empírica, que deve começar com fatos objetivos, ou seja, sobre os quais deve haver acordo intersubjetivo, e que o acesso a eles se dá pela observação. A autora apresenta sua concepção de "fato objetivo" como "aquele que é dito numa determinada linguagem, construída para dar conta das relações supostas pela teoria" (Prado, 1988, p.411), argumentando que eventos observados só têm sentido quando compreendidos no contexto de uma teoria. A autora se pergunta, a seguir, sobre qual seria o ponto de partida de uma observação que se propusesse não ser guiada por nenhuma teoria. E considera que tal empreendimento não seria possível, que o "dado puro" é uma ilusão, e que todo pesquisador se baseia em uma "versão de mundo", seja ela científica ou não e tenha sido explicitada ou não.

Uma consciência crescente da necessidade de vinculação da coleta de dados a quadros de referência teóricos levou os pesquisadores a aprimorarem a vinculação de seus estudos aos modelos teóricos existentes ou a buscarem modelos teóricos alternativos. Essas buscas são apresentadas a seguir.

O estudo da interação social: considerações teóricas e metodológicas

Grande parte dos estudos recentes de observação têm sido realizados em situação de interação social, abordando a interação màe-**criança** (Dessen, Mettel, 1984; Linhares, Marturano, 1984a, 1984b; Rossetti-Ferreira, 1986; Lyra, 1988; Ramos Cerqueira, 1988a, 1988b; Batista, 1989), a **interação professor-aluno** (Machado, 1979; Morais, 1980, 1987; Duran, 1981, 1987; Marturano, Bertoldo, Camelo, 1982; Marturano, 1984, 1986, 1987; Marturano, Maimoni, 1989; Linhares, 1988; Gil-Ceneviva, 1985; Gil, 1990, 1991), as **interações entre crianças** (Bussab, 1988; Oliveira, 1988a, 1988b; Vieira, 1988a, 1988b; Pedrosa, 1989; Carvalho, Beraldo, 1989; Carvalho, 1991, 1992; Pedrosa, Carvalho, 1991; Pedrosa, 1992), o **comportamento pró-social nas interações criança-criança e** adulto-criança (Branco, Mettel, 1984; Branco, 1989, 1991; Bastos, Carvalho, 1991; Bussab, 1991; Carvalho, 1991b; Melo, Branco, 1992) e **Interações em atendimento pediátrico** (Crepaldi, Zannon, 1989; Coutinho, Zannon, 1991).

Esses estudos, e muitos outros, têm levado à reflexão sobre fundamentação teórica e propostas metodológicas na pesquisa da interação social. Serão apresentados no presente item, bem como nos itens seguintes, alguns dos desenvolvimentos recentes na área.

O conceito de interação social, segundo Carvalho

Carvalho considera que interação é o que ocorre entre indivíduos. Para essa autora, "a essência do conceito de interação é o de influência ou regulação recíproca, ou seja: cada um, ou a ação de cada um, é diferente, pelo fato de se dar com o outro, do que seria isoladamente (e/ou com um terceiro?); não se explica pelo que cada um é (ou faz), mas por seus efeitos recíprocos. É evidente, portanto, que, segundo esse conceito, a interação social não é o observável. O que se observa são os comportamentos/ações/estados dos indivíduos (ou eventos) em interação. Interação é um estado ou processo (hipotetizado) de regulação recíproca, inferido a partir do comportamento dos interagentes" (Carvalho, 1988, p.512).

A autora discute que, para alguns autores, o termo interação só deve ser usado quando o sujeito alvo responde ao comportamento do sujeito iniciador, pois, caso contrário, não há evidência de que ocorreu interação (efeito de um sobre o outro). A autora considera que, nesse caso, é necessário definir um intervalo temporal dentro do qual a ocorrência de

uma resposta ainda será tomada como critério para se falar em interação, sendo que a duração desse intervalo é arbitrária e, na maioria dos estudos, bastante curta. Sua crítica à adoção desse critério é de que ele reflete a suposição de que o processo de regulação sempre ocorre de forma imediata ou instantânea, o que considera pouco provável. Para ela, a contingência temporal entre respostas não é o único nem o principal indicador de ocorrência de regulação mútua e pode nem mesmo ser um indicador necessário. Um outro argumento da autora contra o uso desse critério (contato reciprocado como unidade mínima para a análise da interação) é o de que está implícita em sua adoção a hipótese de que a regulação entre os parceiros se processa basicamente através de efeitos mútuos de seus comportamentos. Cita outras formas de regulação que podem ser hipotetizadas: a identidade do parceiro, suas características físicas, etc. Para ela, a ausência de respostas explícitas não significa necessariamente inexistência de interação, entendida no sentido teórico como processo de regulação. Concluindo, a autora afirma que está implícita nessa discussão a necessidade de estabelecer distinções entre dois usos do termo "interação": como categoria teórica e como categoria descritiva, sendo que, nesse último caso, qualquer critério de definição é arbitrário e igualmente aceitável, desde que bem explicitado, tendo, porém, implicações para a compreensão teórica do fenômeno abordado.

O estudo dos relacionamentos interpessoais, segundo Hinde

Hinde propõe a criação de uma ciência dos relacionamentos interpessoais. O autor afirma que "estudos dos relacionamentos interpessoais devem levar em conta não apenas o fenômeno em níveis de complexidade crescente (respectivamente, comportamento social, interações sociais e relacionamentos sociais), cada um com aspectos comportamentais e afetivo-cognitivos, mas também dois tipos de dialética: de um lado, a dialética que se estabelece entre relacionamentos sociais e as personalidades dos indivíduos participantes, e, de outro, a dialética entre esses relacionamentos e as várias influências sociais que determinam sua natureza" (Hinde, 1981, p.4). Cada um desses aspectos será analisado a seguir.

No que se refere aos diferentes níveis de análise do fenômeno, o autor considera como **interação** a ocorrência de uma ou mais trocas do tipo: A faz x para B e B faz y em retorno para A. O termo **relacionamento** é reservado para casos em que uma interação é afetada por interações passadas ou pode afetar interações futuras. A descrição dos aspectos comportamentais de uma **interação** deve especificar seu conteúdo (o que os participantes estão fazendo) e a qualidade da interação (incluindo aspectos relacionais como sincronia, existência de objetivo comum, etc). Já a descrição dos aspectos comportamentais de um **relacionamento** deve abranger tanto o conteúdo e a qualidade das interações componen-

tes como, também, a freqüência absoluta e relativa e a distribuição no tempo dessas interações. Assim, observam-se propriedades emergentes à medida que se progride nos níveis de análise, principalmente do nível das interações para os dos "relacionamentos".

Além dos aspectos comportamentais de interações e relacionamentos, Hinde considera que devem ser abordados seus aspectos cognitivos e afetivos, baseando-se no argumento de que muitas vezes o que ocorreu numa interação é menos importante do que aquilo que os participantes pensam que ocorreu, como eles pensam ter se sentido e como o parceiro foi percebido. Para o autor, o curso futuro do relacionamento depende das avaliações que os indivíduos fazem das interações e relações de que participam, em seus aspectos comportamentais, cognitivos e afetivos.

Quanto aos aspectos dialéticos, Hinde menciona em primeiro lugar o fato de que relacionamentos influenciam os indivíduos e que a personalidade desses indivíduos influencia os relacionamentos em que eles se inserem, criticando o fato de que a Psicologia tem devotado muito mais esforços ao estudo dos aspectos individuais da personalidade do que ao estudo dos relacionamentos. O segundo tipo de dialética surge do fato de que cada relacionamento existe em um ambiente social, o que destaca dois aspectos: por um lado, ambos os participantes de um relacionamento influenciam e são influenciados pelo relacionamento em foco; de outro lado, esse relacionamento é afetado pelas normas da cultura em que eles vivem, ao mesmo tempo em que essas normas são alteradas ao longo do tempo pelos relacionamentos diádicos e pelos de ordem mais elevada que constituem a sociedade. Assim, é necessário levar em conta esses dois tipos de dialética: entre personalidade e relacionamento, bem como entre relacionamentos e sociedade.

Tipos de recorte do fluxo interativo, segundo Marturano

Entre as soluções para um estudo da interação social que levem em conta a complexidade dos eventos envolvidos está a distinção entre **tipos de recorte do fluxo Interativo,** apresentada por Marturano (1987). Ao discutir estudos de interação professor-aluno, a autora examina quatro formas de recorte do fluxo interativo, que representam niveis progressivamente mais altos de complexidade e que são aplicáveis a outros tipos de interação, além da interação professor-aluno.

O primeiro procedimento é o que apreende **comportamentos isolados da professora e do aluno,** permitindo levantamentos de freqüência e limitado a estudos correlacionáis.

O segundo refere-se à **transição entre comportamentos da professora e do aluno.** Segundo a autora, as posições relativas das respostas no fluxo temporal são tomadas como referência para a cons-

trução de matrizes de dupla entrada, tendo como antecedentes ora os comportamentos do aluno, ora os do professor. Podem-se obter, assim, informações de âmbito limitado sobre relações de reciprocidade e controle.

O terceiro procedimento propõe a divisão do fluxo interativo em **episódios,** ou seja, segmentos delimitados por condições observáveis de início, desenvolvimento e término. Além das análises realizadas pelo segundo tipo de procedimento, é possível também investigar separadamente as iniciativas do aluno e do professor. A autora destaca que esse recorte permite a derivação de indicadores quantitativos de reciprocidade e controle, bem como de centralidade e assimetria na organização do grupo.

A quarta modalidade de recorte localiza as **relações entre episódios adjacentes,** levando em conta diferentes possibilidades de transição, superposição ou interseção entre episódios, permitindo conceber a sala de aula como grupo centralizado na professora e não apenas como conjunto de episódios diádicos isolados, como é o caso do terceiro tipo de recorte.

Uma representação esquemática dos quatro tipos de recorte, bastante elucidativa, é apresentada no texto original (Marturano, 1987).

Categorías intra e inte/individuáis na análise de interações, segundo Carvalho

Carvalho (1988) considera que, se a interação é um evento que ocorre entre indivíduos, conforme discutido anteriormente, esta não poderia ser adequadamente descrita com base em categorias referentes ao comportamento de um dos parceiros, definidas independentemente do comportamento do outro. O tipo de categoria adequado seria aquele cuja definição envolvesse a consideração de ambos os parceiros e/ou de relações entre seus comportamentos.

A autora faz a distinção entre sistemas de categorias baseadas numa **dimensão estrutural** da interação, dos baseados nos **conteúdos.** Entre as ligadas à **dimensão estrutural,** estão as que se fundam em relações de qualquer tipo entre os comportamentos dos parceiros, por exemplo, relações de semelhança, relações temporais, etc; esse é o caso de categorias como "interação complementar", "interação simétrica", etc. Já entre as categorias baseadas no "conteúdo", a autora menciona exemplos que considera claramente interindividuais, como "imitação", "brinquedo paralelo", "cooperação", bem como outros em que nem sempre é claro se a categoria se refere ao comportamento de um indivíduo emissor ou se depende da consideração de dois ou mais parceiros, como é o caso de "agressão", "ajuda", etc.

Sistemas de categorias para análise do comportamento verbal, segundo Marturano

Marturano (1988) tece considerações sobre a lógica do sistema de categorias, aplicada à análise do comportamento verbal. Afirma que, na

abordagem do comportamento verbal através de sistemas de categorias, surgem dificuldades na elaboração e na utilização do instrumento, que dizem respeito principalmente aos requisitos de exclusão mútua e coerência de critérios. Em relação ao primeiro requisito (ou seja, a necessidade de que as categorias sejam mutuamente exclusivas), a autora afirma que, algumas vezes, na prática, a inclusão de um segmento de fala em uma ou outra categoria é muito mais uma questão de decisão do categorizador que uma decorrência lógica das respectivas definições. E com relação ao segundo requisito (ou seja, de que haja coerência na utilização dos critérios de categorização), considera que as distorções decorrem em geral de duas ordens de problema: a) categorias distintas são definidas a partir de critérios diferentes; b) na aplicação do instrumento, utilizam-se pistas adicionais não previstas nas definições e decorrentes do conhecimento prévio do categorizador como usuário da linguagem.

A autora apresenta uma concepção do comportamento verbal como "multidimensional, integrado a um sistema de comportamento mais amplo e constituindo ele próprio uma cadeia de eventos interdependentes" (Marturano, 1988, p.518). Para dar conta de toda essa complexidade, ela sugere que se tomem os seguintes cuidados, quando diferentes dimensões do comportamento verbal forem tratadas em um único sistema de categorias: a) explicitar as dimensões cuja inclusão na análise o problema de pesquisa requer; b) explicitar os critérios para identificação de variações em cada dimensão; c) incluir em todas as definições de categorias a referência ao *status* de cada dimensão e os respectivos indicadores, de acordo com os critérios previamente estabelecidos.

Como alternativa, sugere a utilização de sistemas múltiplos, ou seja, um sistema de categorias independentes para cada dimensão focalizada. Nesse caso, são feitas várias leituras dos dados, e cada unidade de verbalização recebe tantas classificações quantas sejam as dimensões estudadas. A autora apresenta como uma das vantagens do procedimento de sistemas múltiplos o fato de permitir uma resolução qualitativa mais diferenciada no tratamento dos dados, uma vez que podem fornecer, através da análise de dependência, configurações formadas pela associação significativa entre categorias de diferentes sistemas; dessa forma, as relações entre dimensões são extraídas empiricamente dos dados. A autora aponta o trabalho de Ramos (1979) como um exemplo da realização da proposta de leituras múltiplas do comportamento verbal.

A construção de Unidades de Análise Comportamental (UAC) para o estudo do comportamento pró-social, segundo Branco e Mettel

Branco (1989) e Branco e Mettel (1989a, 1989b), visando ao estudo do comportamento pró-social, realizaram gravações em vídeo e vários outros tipos de registro em uma sala de maternal, com crianças cuja idade estava em torno de cinco anos. Branco (1989) relata que, para transcrição

do vídeo, deu preferência à transcrição exclusiva dos episódios considerados relevantes, ao invés da transcrição detalhada da fita toda. Para tanto, o mesmo trecho da fita era observado repetidamente, até que fossem identificados os episódios que pareciam apresentar alguma relação com interações criança-criança e definidos os critérios de identificação desses episódios.

Em seguida, os episódios foram decompostos em unidades de análise, compreendendo três elementos básicos: um segmento do comportamento da professora, uma situação antecedente especificando a natureza da relação entre as crianças com quem a professora interagiu e uma situação conseqüente especificando a ocorrência ou não de alterações referentes à relação criança-criança descrita como antecedente. Cada unidade devia referir-se a um único padrão criança-criança, excetuando-se aquelas que representassem verbalizações da professora que não definiam um padrão criança-criança específico. Dessa forma, cada um dos padrões criança-criança que se relacionavam no episódio ao comportamento da professora geraram uma Unidade de Análise Comportamental (UAC). Uma vez determinadas as UAC, passou-se à sua classificação através do desenvolvimento de dois sistemas de categorias articulados, relacionando ações da professora aos diversos padrões criança-criança. "Cada UAC é, portanto, composta por uma categoria correspondente à ação da professora e outra ao padrão criança-criança ao qual ela faz referência ou sobre o qual ela atua" (Branco, 1989, p.78). As ações da professora, num total de 14, foram categorizadas em três classes: estabelecimento, supressão e verbalização de regras, sempre referentes a algum tipo de categoria de relacionamento criança-criança. E as 4ª categorias específicas do relacionamento criança-criança foram classificadas em quatro tipos diferentes: atenção ao colega, interações sintónicas, interações não-sintônicas e competição planejada. Além disso, as UAC foram identificadas em cada uma das 12 categorias relativas às atividades desenvolvidas na sala de aula. Das 488 UAC registradas, 419 foram contingentes à ocorrência de algum tipo de relacionamento criança-criança, sendo que, destas, 76% relacionaram-se ao estabelecimento, 17% à supressão e 7% à verbalização de regras relacionadas aos diferentes padrões criança-criança. Verificou-se, também, que a situação de roda foi o principal contexto para a observação de UAC.

Um exemplo de análise funcional da interação professor-aluno, segundo Gil

Gil (1990, 1991) estudou as relações entre o desempenho de uma professora de primeira série do primeiro grau e o desempenho de seus alunos, durante a realização de uma rotina de sala de aula, composta por seqüências comportamentais simples e curtas (por exemplo: seqüências rítmicas de batidas de palma ou de lápis na mesa, seqüências de movi-

mentos corporais, etc.). Levando em conta que deveriam ocorrer ajustamentos recíprocos e contínuos nos repertórios da professora e dos alunos, a autora buscou trabalhar também com estados transicionais, além das classes de desempenho e sua freqüência. Nesse sentido, "as definições das classes foram amplas e/ou provisionais, isto é, foram classes funcionais que mudavam à medida que suas funções, estabelecendo-se, ensejavam novas funções. Assim, a necessidade de trabalhar com fluxos de desempenho da professora e do aluno gerou a necessidade de registrarem-se produtos específicos destes fluxos. Estes, algumas vezes, seriam mudanças no ambiente físico do outro, mas, mais freqüentemente, seriam mudanças no próprio repertório estudado, vistas não só como uma classe, mas como uma série de comportamentos literalmente seriados" (Gil, 1991, p.315).

A autora lembra que para fazer uma análise funcional dos desempenhos da professora e dos alunos deveria medir o efeito de um evento sobre o outro. Dessa forma, estabeleceu recortes nos momentos em que produtos específicos puderam ser identificados. No fluxo de desempenho dos alunos, foram considerados produtos os exemplos de desempenhos recorrentes, posteriores aos desempenhos recorrentes da professora. E no da professora, foram considerados produtos os exemplos recorrentes de seu desempenho, posteriores aos desempenhos recorrentes dos alunos. A autora considera que os limites dessa recorrência apresentaram-se como um problema tanto empírico quanto conceptual.

A coleta de dados foi realizada através de gravações em vídeo e dos registros transcritos, levando-se em conta as emissões vocais da professora e dos alunos e parâmetros não-verbais de seus desempenhos. Examinando-se a transcrição dos dados, foi selecionada para análise uma das rotinas de sala de aula, caracterizada pelo fato de os alunos reproduzirem de maneira formal e/ou literal modelos sonoros e/ou gestuais apresentados pela professora, denominando-se essa rotina de "seqüência". A autora listou e descreveu as ocorrências das seqüências, cada uma das quais sendo considerada um episódio.

Foi preenchido, a seguir, um protocolo, contendo descrição dos desempenhos da professora e dos alunos durante as ocorrências da rotina (seqüência) e durante as situações imediatamente anteriores e posteriores a estas ocorrências. Indicou-se, ainda, a natureza das atividades desenvolvidas em classe antes e depois das seqüências. A análise desses protocolos permitiu identificar: a) quatro classes ou tipos de seqüências (seqüência rítmica de sons e movimentos, seqüência de palavras ou frases e movimentos, seqüência de instruções e movimentos, seqüência de movimentos); b) três classes de atividades em sala (organização da classe, ensino-aprendizagem de requisitos para leitura e escrita, ensino-aprendizagem de leitura/escrita); c) duas classes de participação dos alunos (cumprir tarefa, dispersão); d) seis classes para classificação do

tipo de barulho produzido pelos alunos (falas isoladas, falas em uníssono, burburinho, gritaria, riso e silêncio). As classes apresentadas nos itens b, e e d foram utilizadas para analisar as situações anteriores e posteriores à apresentação das seqüências.

Foi analisada a freqüência de ocorrência das seqüências e sua relação com as atividades de classe anteriores e posteriores, considerando-se relações de dependência através de probabilidades condicionais. Essa análise indicou relações de dependência entre as ações da professora e as dos alunos (por exemplo: uma "seqüência rítmica de sons e movimentos" (SMR) tinha o efeito de reduzir totalmente "riso" e "gritaria" e de reduzir acentuadamente "burburinho"; tinha, também, o efeito de produzir um aumento de "cumprir tarefa" e uma redução de "dispersão". Por sua vez, os antecedentes de ruído de classe mais freqüentes das seqüências SMR eram "burburinho", "gritaria" e "riso", e nunca "silêncio", "falas isoladas" ou "falas em uníssono". Dessa forma, os dados analisados indicam "que o desempenho da professora e dos alunos são mútua e finamente modulados" (Gil, 1991, p.318).

A perspectiva sociointeracionlsta-construtivista de estudo do desenvolvimento

Trata-se de uma perspectiva que aborda a situação de interação social como constitutiva dos sujeitos. Enfoca o diálogo como troca negociada constitutiva dos sujeitos e da própria troca (Lyra, 1988). Baseia-se, em parte, em formulações referentes à aquisição da linguagem (Lemos, 1981), em que a autora considera que as situações de interação são um espaço para a construção de relações semânticas, morfológicas e sintáticas pela criança. Ela propõe uma continuidade entre o período pré-verbal e o verbal na aquisição da linguagem e sugere que a criança deve passar pelo processo de inserir-se numa estrutura diàdica com um interlocutor para construir uma representação de si mesma e do outro como interlocutores. Considera que sua proposição é semelhante à de Vygotsky, citado por ela, de que as atividades da criança adquirem significado em um sistema de relacionamento social, na medida em que o caminho que vai do objeto para a criança e da criança para o objeto passa pela relação com outra pessoa.

Dentro desse enfoque, Lyra e colaboradores (Lyra, 1988,1991; Lyra, Galindo, Cipriano, 1990; Lyra, Cabral, Pantoja, 1991) têm desenvolvido um programa de pesquisa das interações entre mãe e bebê, realizando registros longitudinais em vídeo em situação natural. Têm sido analisadas como atividades partilhadas as interações "face a face" e as interações "mãe-objeto-bebê". Nas interações "face a face", têm sido estudadas a produção vocal e o sorriso. Ao discutir aspectos referentes à transcrição e análise de dados em vídeo, Lyra (1991) afirma a necessidade de uma "filiação profunda" entre a posição teórico-epistemológica do pesquisador

e a metodologia de transcrição e análise de dados a ser adotada. Esclarece alguns pontos acerca de sua concepção de estudo, destacando-se: a) a necessidade de exploração de registros longitudinais; b) uma concepção de causalidade como transformações e construções bidirecionais e interdependentes; c) o enfoque em ambos os sujeitos bem como na interação entre eles; d) a visão da diade como unidade de análise indissociável. Nesse sentido, propõe como solução metodológica para uma primeira fase do programa de pesquisa a descrição e análise de momentos identificados como qualitativamente diferentes dos anteriores e posteriores, "procurando relacionar as transformações e construções identificadas à compreensão do dinamismo existente na história da diade, ou seja, aos processos de troca negociada entre os parceiros, chamados de processos dialógicos" (Lyra, 1991, p.294). Decorre daí a necessidade do registro longitudinal em vídeo, cuja transcrição e análise são efetuadas a partir de critérios derivados dessas concepções.

A partir dessa análise, de cunho qualitativo, a autora relata o destaque de um elemento presente nas organizações face a face: as trocas vocais, que ensejaram uma segunda fase de seu programa de pesquisa, incluindo uma verificação quantitativa das modificações identificadas na produção vocal do bebê. Reafirma, entretanto, que sua posição teórica vê a quantidade como filiada à qualidade, sendo que a compreensão do processo de transformação e construção continuam a depender de uma análise qualitativa.

Outro exemplo de trabalho nessa perspectiva é o de Oliveira (1988a), que realizou um estudo longitudinal em uma creche pública, usando a noção de "jogo de papéis" como foco da análise dos dados, registrados em vídeo. A análise das transcrições foi realizada com base na seleção de trechos em que se evidenciava a construção de enredos de diferentes papéis construídos pelas crianças no grupo e no faz-de-conta. A análise mostrou a construção coletiva da brincadeira e uma crescente diferenciação de papéis. A autora tem analisado também situações de interação adulto-criança em creches e pré-escolas. Ela discute o conceito de papel na análise do processo interacional, retomando postulados da Psicologia sociointeracionista de base dialética do início do século, e propõe um enfoque processual que concebe a interação "como constituindo os sujeitos e não como influências, mesmo que recíprocas, entre sujeitos já constituídos" (Oliveira, 1992).

Ainda um outro exemplo de pesquisa abordando a interação entre crianças como um processo de construção mútua é o trabalho de Pedrosa (1989), que fez gravações em vídeo na creche, em situação de recreação livre. A autora apresenta seus dados na forma de descrição de onze episódios, indicando para cada um: a) o nome do episódio; b) dados sobre a data, horário, duração, crianças envolvidas; c) descrição da situação (presente e imediatamente anterior); d) resumo do episódio e descrição

detalhada do mesmo. A análise dos dados evidencia: a) a ocorrência de brincadeira compartilhada em grupo, sem a sugestão do adulto, em crianças desde menos de dois anos de idade; b) o levantamento de hipóteses sobre processos de regulação, tais como o "arranjo" da brincadeira, ajustamentos rítmicos e posturais, regras, etc. Discutindo questões referentes à transcrição e análise de registro em vídeo, Pedrosa e Carvalho (1991, p.285) definem episódio interativo como "um segmento de registro onde aparecem seqüências interativas claras e conspicuas ou trechos do registro em que se pode circunscrever um grupo de crianças a partir do arranjo que formam e/ou da atividade que realizam" e consideram que o recorte em episódios corresponde a uma primeira forma de seleção dos dados que parece útil para perseguir a estrutura interacional do grupo. Afirmam que o início de um episódio pode ser delimitado por um fato a partir do qual se articula uma seqüência interativa (por exemplo: uma criança pega um apito e começa a soprá-lo) ou o episódio pode emergir de uma situação em que vários elementos se combinam e se constituem em uma seqüência, às vezes só reconhecida a *posteriori,* e cujo início é algumas vezes arbitrado, o mesmo ocorrendo para a delimitação do término de um episódio.

Os episódios são a seguir divididos em momentos, com critérios bastante flexíveis. Em alguns casos, segue-se a composição social do grupo; em outros, a configuração da seqüência em curso, seguindo pistas como o conteúdo da brincadeira, a distribuição espacial, etc. As autoras ressaltam que essa proposta de recorte decorre de sua concepção de interação, concebida como "um processo de influências recíprocas entre os parceiros" (id. ibid., p.286), admitindo que a estruturação desses arranjos envolve processos de regulação entre as crianças.

As autoras destacam dois tipos de transcrição do episódio: um que privilegia uma seqüência de interação identificada e descreve o comportamento das outras crianças que também estão em cena, apenas se estiver diretamente relacionado àquela seqüência — nesse caso, o interesse é seguir a trama interacional específica; já o outro tipo de transcrição procura incluir tanto as seqüências interacionais quanto informar sobre todos os comportamentos das crianças que aparecem em cena, com um foco de atenção mais abrangente, usado quando se deseja comentar as regulações recíprocas entre as crianças presentes num registro.

Quanto ao nível de detalhamento na transcrição do episódio, Pedrosa e Carvalho propõem como regra: o comportamento das crianças deverá ser transcrito em detalhes apenas se essa descrição for esclarecedora para a compreensão do que se deseja demonstrar. Com base nesta regra, só se descreve, por exemplo, a postura da criança em suas minúcias morfológicas quando esta for o alvo de interesse do trabalho (por exemplo: uma inclinação de tronco da criança pode interessar quando for seguida de movimento semelhante por outra criança).

As autoras sugerem, ainda, o registro de impressões provocadas no pesquisador pelas ações das crianças, distinguindo-as dos fatos observados, dando um "tom" à situação e favorecendo a compreensão do episódio.

A análise qualitativa é proposta a partir de perguntas que levam ao exame e ao reexame dos episódios e à sua relação com as formulações teóricas subjacentes.

Assim, dentro da perspectiva sociointeracionista-construtivista, têm sido exploradas novas formas de análise e tratamento de dados, decorrentes das formulações teóricas adotadas pelos pesquisadores da área.

A perspectiva ecológica de estudo dos arranjos ambientais

A perspectiva ecológica de análise vem se delineando dentro da Psicologia ambiental que, segundo Campos de Carvalho (1988, p.305), "se propõe a estudar as relações entre ambiente e comportamento humano, enfatizando a influência do ambiente físico no comportamento. Pela abordagem ecológica, todo e qualquer contexto ambiental é visto como um sistema de inter-relações, ou de interdependência, entre os vários componentes físicos e humanos que participam daquele contexto. Somente para efeito de estudo e análise é que se abstrai deste sistema dinâmico um componente; mas ele só pode ser compreendido em relação àquela situação da qual foi extraído. Cada componente é tanto causa como efeito — ele atua sobre os outros componentes assim como esses atuam sobre ele".

Dentro dessa perspectiva, a autora realizou estudos em duas creches que atendem crianças de nível sócio-econômico baixo, na região de Ribeirão Preto, replicando estudos de Legendre sobre arranjos espaciais em creches.

Legendre, citado pela autora, observou três tipos de arranjos espaciais em creches: a) **Arranjo espacial semi-aberto** — caracterizado pela presença de zonas circunscritas (ZC), que são áreas delimitadas pelo menos em três lados por barreiras formadas por elementos do mobiliário, parede, desnível do solo, etc. As crianças em geral ocupam as ZC, ficando afastadas do adulto, porém com fácil acesso a ele, inclusive visual; observa-se, nesse caso, um favorecimento da ocorrência de interações afiliativas entre crianças e um tipo de interação menos freqüente e mais intenso com o adulto (ou seja, evocando mais respostas deste), b) **Arranjo espacial aberto** — caracterizado pela ausência de ZC, geralmente havendo um espaço central vazio. Neste arranjo, decrescem as interações entre crianças, as quais tendem a permanecer próximas ao adulto, embora com pouca interação com ele. c) **Arranjo fechado** — caracterizado pela presença de ZC, porém havendo barreiras físicas (por exemplo, um móvel alto) dividindo a sala em duas ou mais áreas. Assim, as barreiras físicas impedem uma fácil visão do adulto, embora permitam a circulação das crianças pela sala. Neste arranjo, as crianças tendem a permanecer em volta do adulto, ocorrendo pouca interação entre as crianças.

Campos de Carvalho (1988, 1990) replicou esse estudo em duas creches, caracterizadas por uma proporção relativamente grande de crianças por adulto, geralmente em espaços vazios e com pouca disponibilidade de objetos. Uma vez que este tipo de ambiente não favorece a interação criança-adulto, nem a interação entre crianças com idade inferior a três anos, buscou-se verificar de que forma condições ambientais favoráveis agiriam como suportes ou mediadores na interação entre crianças pequenas. Nesse estudo, foi feita uma transformação do arranjo espacial aberto para semi-aberto, na área utilizada pelas crianças da creche para atividades livres. Os dados foram coletados nesses períodos, por duas câmaras fotográficas, que funcionavam automaticamente e em conjunto a cada 30 segundos, anotando-se, posteriormente, a posição de cada criança e da pajem em cada foto. Os dados obtidos foram na mesma direção dos de Legendre: a estruturação do espaço propiciou oportunidades de contato entre as crianças, permitindo, ao mesmo tempo, que o adulto estivesse mais atento ao grupo de crianças. Esses aspectos foram aproveitados no projeto de intervenção, realizado pela autora após a coleta dos dados desta pesquisa.

Estudos posteriores, utilizando gravação em vídeo, têm permitido a identificação de diferenças no comportamento das crianças, quando se compara o que ocorre na área em torno da pajem com o que ocorre nas zonas circunscritas, dando prosseguimento a essa linha de trabalho (Rubiano, 1990, 1991; Chaguri, Rubiano, Silva, Rossetti-Ferreira, 1991).

O estudo dos movimentos expressivos

Nos primeiros estudos observacionais, era freqüente a orientação para que o observador se ativesse aos elementos observáveis e definíveis em termos físicos, evitando-se inferências e impressões subjetivas. Em certos casos, essas recomendações levaram a uma deturpação, no sentido de se evitar a abordagem a certos elementos sutis do comportamento, tais como expressões faciais de curta duração e difíceis de definir (como, por exemplo, uma expressão de nojo rapidamente substituída por uma expressão mais neutra), posturas corporais que produzem impressão de dominação ou submissão, verbalizações interpretáveis como "irônicas", etc. Entretanto uma área de pesquisa designada por comunicação não-verbal (Scherer, Ekman, 1982; Corraze, 1982), que tem abordado a face, o olhar, a voz (entonação, ritmo, etc), a postura, o movimento corporal, o olfato e o tato como meios de comunicação. São empregados dois métodos básicos de investigação: **o estudo através dos julgamentos** e **o estudo através dos constituintes.** Corraze (1982, p.69) descreve esses métodos no contexto do estudo das expressões faciais. Ele diz que **"partir dos julgamentos** consiste em apoiar-se em múltiplas expressões faciais que se apresentem a observadores e em procurar saber se, entre eles, há concordância quanto à sua significação afetiva. A face é então considerada

corno um estímulo suscetível de produzir uma resposta, um julgamento, em observadores". O autor considera que o estudo pelos julgamentos pode nos dizer que a informação está ali, mas não nos pode dizer que comportamentos da face levam a essa informação precisa. Já **o estudo pelos constituintes** tenta identificar os componentes da face que correspondem à produção de determinadas emoções nos sujeitos que fizeram os julgamentos. Esses dois métodos são extensivos ao estudo de outros aspectos da comunicação não-verbal, além das expressões faciais.

Um estudo de julgamento de expressões faciais foi realizado por Silva (1987a), utilizando estudantes universitários como sujeitos. O autor verificou que os julgamentos das expressões faciais, mostradas através da projeção de diapositivos, apresentavam fidedignidade intra-sujeito, entre grupos e intercultural. O autor também apresentou evidências de que os erros cometidos nos julgamentos de expressões faciais de emoções não eram aleatórios, sendo os erros dos brasileiros praticamente idênticos aos cometidos pelos sujeitos americanos.

Silva (1987b) destaca três questões principais que têm norteado a maioria das pesquisas sobre expressões faciais de emoções. A primeira diz respeito à existência de expressões faciais características dos diferentes estados emocionais, a segunda, à universalidade dessas expressões, e a terceira, aos sinais faciais típicos de cada uma dessas expressões. O autor considera que as três questões foram respondidas positivamente, a partir das pesquisas realizadas pelas equipes de Paul Ekman e de Carroll Izard. Considera, também, que restam várias questões a serem melhor investigadas, entre as quais se incluem as seguintes: a) que emoções possuem expressões faciais mais claramente reconhecíveis? b) como distinguir as expressões faciais referentes a emoções genuínas daquelas que aparecem devido à mentira facial ou devido a outros motivos? c) como são os padrões temporais das expressões faciais de emoções? d) qual a unidade "natural" de julgamento? (a emoção do outro, em que condições, ou qualidades mais permanentes de quem sente tal emoção?) e) em uma situação "natural", como a pessoa julga expressões faciais de emoção? f) qual a proeminência das expressões faciais de emoções em relação a outras informações, sobre emoções ou não, presentes em uma dada situação sob avaliação? O autor considera que os estudos a serem propostos por estas questões provavelmente irão envolver outras estratégias de pesquisa, além da análise de expressões faciais em fotografias.

As expressões faciais também têm sido utilizadas como indicadores de paladar e olfato em recém-nascidos, em estudos da ontogenèse da quimiorrecepção (Bergamasco, 1991; Bergamasco, Lima, Krusnaukas, 1991; Bergamasco, 1992). O procedimento de registro consiste em filmar em vídeo as respostas de recém-nascidos antes, durante e após a apre-

sentação de cada estímulo, gustativo ou olfativo. O material obtido é trabalhado por observadores não informados sobre o estímulo apresentado ao bebê, através de uma ou mais das seguintes formas: levantamento de categorias, análise temporal da resposta e avaliação por juizes. Os dados obtidos têm permitido a identificação de respostas diferenciadas e características para diferentes tipos de estímulo.

Um outro tipo de estudo que tem sido realizado é o do sorriso, em seus aspectos morfológicos e motivacionais (Fagundes, 1978; Otta, Sarra, 1988), como reação do bebê a diferentes estímulos (Obara, Akamine, Pedrazzoli Neto, Bonilha, Bortoletto, Otta, 1990) e como influência na percepção de pessoas, quando se variam as características do sujeito e do modelo (Delevati, Lira, César, Pires, Otta, 1992).

Através destes e de muitos outros estudos, a área de estudo dos movimentos expressivos tem permitido a identificação e análise de muitos dos elementos sutis que levam à formação de impressões subjetivas ao longo das relações interpessoais.

A pesquisa participante e a observação antropológica

A pesquisa participante tem sido proposta por sociólogos, antropólogos e educadores, como um meio de superar a dicotomia e o distanciamento entre "sujeito" e "objeto" de pesquisa, de forma que pessoas, anteriormente "objetos de pesquisa", passem a ser "o **sujeito,** tanto do **ato** de **conhecer** de que têm sido **o objeto,** quanto do **trabalho de transformar** o conhecimento e o mundo que os transformaram em objetos" (Brandão, 1981, p.11).

Thiollent (1984) estabelece uma distinção segundo a qual a "pesquisa participante" preocupa-se principalmente com o papel do investigador dentro da situação investigada, enquanto a "pesquisa-ação" é uma modalidade de "pesquisa participante" centrada em projetos de ação social ou de resolução de problemas coletivos. O autor afirma que a "pesquisa participante" tem sido concebida como uma técnica de "observação participante", elaborada no contexto da pesquisa antropológica ou etnográfica. Segundo ele, "trata-se de estabelecer uma adequada participação dos pesquisadores dentro dos grupos observados de modo a reduzir a estranheza recíproca. Os pesquisadores são levados a compartilhar, pelo menos superficialmente, os papéis e os hábitos dos grupos observados para estarem em condição de observar fatos, situações e comportamentos que não ocorreriam ou que seriam alterados na presença de estranhos" (Thiollent, 1984, p.83).

A observação "antropológica" tem sido considerada por alguns autores como uma alternativa recomendável para o estudo de diferentes situações sociais. É o caso, por exemplo, de Delamont, que critica o uso

de sistemas de categorías preestabelecidas, como, por exemplo, o FIAC *(Flanders Interaction Analysis Categories),* para o estudo da interação professor-aluno (Delamont, Hamilton, 1981; Patto, 1981). Delamont e Hamilton (1981) consideram que a análise antropológica se vale de uma abordagem mais etnográfica do que "psicométrica" e de uma estrutura conceptual que considera a educação em termos socioculturais mais amplos. Metodologicamente, esses estudos baseiam-se na observação participante, que abrange a presença de um ou mais observadores durante longos períodos, numa única sala de aula ou num pequeno número delas. Durante esse tempo, o etnòlogo também conversa com os participantes, referidos como "informantes", ao invés de "sujeitos". Ele pode conduzir entrevistas formais e também aplicar questionários. Esses autores consideram que o "antropólogo" tem um quadro de referência holístico, e que não procura manipular, controlar ou eliminar variáveis. Sua seleção de dados se faz através da concentração da atenção em aspectos emergentes, e não através de categorias preestabelecidas.

Patto (1981) comenta um artigo de Delamont, em que esta autora estabelece comparação entre dados coletados através do FIAC com dados obtidos através de uma observação antropológica, conforme acima descrita. Para um caso em que duas professoras de Latim eram semelhantes em relação a algumas categorias de Flanders e diferentes em outras, a autora utilizou observações não-estruturadas e entrevistas formais e informais com as alunas. Para a estruturação dos dados de observação, a autora selecionou alguns temas unificadores, que emergiram das entrevistas com as alunas e das próprias observações e anotações de campo. Entre os temas que se destacaram, incluem-se: o ambiente físico criado pelas professoras (caracterizado, principalmente, pelas anotações de campo), a aparência pessoal das professoras (através de conversas informais com as alunas), as opiniões das alunas sobre as professoras (por meio de entrevistas formais com as alunas) e trechos de diálogos ocorridos durante a aula (através de anotações detalhadas tomadas em sala de aula). A partir desses dados e de considerações sobre a natureza da matéria ensinada, a autora conseguiu entender várias das diferenças entre as duas professoras acima referidas, enquanto o sistema de categorias preestabelecidas apenas constatou algumas delas.

No mesmo texto, Patto cita outro estudo, de Walker e Adelman, que aborda a complexidade dos significados comunicativos em sala de aula, que aparecem nas relações informais que se estabelecem entre o professor e os alunos, em especial nas piadas e brincadeiras que se dão em determinados momentos. Para tanto, foi necessário reconstituir a história da turma, inserindo os diálogos em seu contexto espacial e temporal mais amplo. Entre as técnicas de observação utilizadas por esses autores, incluíram-se: filmagem e gravação das aulas, observações intensivas durante períodos curtos e longos de tempo, acompanhadas de anotações,

consultas às notas dos professores, seus planos de aula, entrevistas com os professores e com os alunos. Patto considera que a percepção do significado das comunicações nesse estudo foi possibilitada principalmente pela pesquisa participante de longa duração, bem como pelas entrevistas informais com professores e alunos. A autora afirma que o uso da observação antropológica pode enriquecer a compreensão das relações entre professor e alunos na sala de aula.

O presente texto assemelha-se mais a um "menu de opções" do que a "manual de instruções". Relata a evolução dos estudos observacionais em nosso meio, a partir dos trabalhos iniciais influenciados pela análise do comportamento e pela etologia animal, e apresenta alguns dos desenvolvimentos recentes na área, na expectativa de que o aprimoramento dos métodos observacionais, utilizados em conjugação com outros métodos de investigação em Psicologia, contribuam para a melhor compreensão do ser humano, especialmente em sua dimensão social.

REFERÊNCIAS BIBLIOGRÁFICAS

ADES, C. *A teia e a caça da aranha Argiope argentata.* São Paulo, 1972. Tese (Doutorado) — Faculdade de Filosofia. Letras e Ciências Humanas, USP.

ALTMANN, J. Observational study of behavior: sampling methods. *Behaviour,* v.49, p.227-267, 1974.

ALTMANN, S.A. Sociobiology of rhesus monkeys: the stochastics of social communication. *Journal of Theoretical Biology,* London, v.8, p.490-522, 1965.

ALVES, Z.M.B.A. *Análise da interação máe-criança:* um estudo longitudinal (dos dois ao seis meses) da evolução de seqüências de interação. Ribeirão Preto, 1973. Tese (Doutorado) — Faculdade de Filosofia, Ciências e Letras, USP.

ALVES, Z.M.B.A., CARVALHO, A.M.A., MARTURANO, EM., METTEL. T.P.L. ROSSETTI-FERREIRA, M.C. Questões metodológicas no estudo da interação humana. In: REUNIÃO ANUAL DA SOCIEDADE DE PSICOLOGIA DE RIBEIRÃO PRETO, 17, 1987. *Anais...* Ribeirão Preto, 1987. p.531-541.

BAER, D.M., WOLF, M.M., RISLEY, T. Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis,* Lawrence (KS), v.1, n.l, p.91-97, 1968.

BASTOS, M.F., CARVALHO, A.M.A. Empatia entre crianças. In: REUNIÃO ANUAL DASOCIEDADE DE PSICOLOGIA DE RIBEIRÃO PRETO, 21,1991. *Anais...* Comportamento pro-social: questões filogenéticas e ontogenéticas. Ribeirão Preto, 1991. p.l 14-116. «

BATISTA CG. *Catálogo de comportamentos motores observados durante uma situação de refeição.* São Paulo, 1978. Dissertação (Mestrado) — Instituto de Psicologia, USP.

--------------. Concordância e fidedignidade na observação. *Psicologia,* v.3, n.2, p.39-49,1977.

--------------. Elaboração de um catálogo de comportamentos motores observados na interação entre pais e filhos. *Psicologia,* v.6, n.3, p.47-81, 1980.

--------------. *Mães e crianças brincando:* um estudo de influências recíprocas. São Paulo, 1989. Tese (Doutorado)— Instituto de Psicologia, USP.

BATISTA, CG. Objetivos da avaliação da fidedignidade em estudos observacionais. *Psicologia: Teoria e Pesquisa,* v.1, n.3. p.205-214, 1985.

BATISTA. CG., MATOS, MA. O acordo entre observadores em situação de registro cursivo: definições e medidas. *Psicologia,* v.10, n.3, p.57-69, 1984.

BERGAMASCO, N.H.P. Expressões faciais de recém-nascidos em resposta a estímulos gustativos e olfativos. In: REUNIÃO ANUAL DA SOCIEDADE DE PSICOLOGIA DE RIBEIRÃO PRETO. 21. 1991. *Anais...* Transcrição e análise de registro em vídeo. Ribeirão Preto. 1991.p.291-292.

. . Percepção de estímulos químicos em recém-nascidos: discriminação e preferências. In: REUNIÃO ANUAL DASOCIEDADE BRASILEIRA DE PSICOLOGIA, 22.1992. *Resumos de comunicações cientificas...,* n.336.1. [S.I.], 1992.

BERGAMASCO, N.H.P, LIMA, A., KRUSNAUKAS, I. Valor de comunicação das respostas faciais de recém-nascidos. In: REUNIÃO ANUAL DA SOCIEDADE DE PSICOLOGIA DE RIBEIRÃO PRETO, 21, 1991. *Resumos...,* n.211. Ribeirão Preto, 1991.

BIJOU, S.W., PETERSON, R.F., AULT, M.H. A method to Integrate descriptive and empirical field studies at the level of data and empirical concepts. *Journal of Applied Behavior Analysis.* Lawrence (KS), v.1, n.2, p.175-191. 1968.

BLURTON JONES, N. (Org.). *Ethological studies of child behavior.* Cambridge: Cambridge University Press, 1972.

BRANCO, A.U. Ontogenèse e promoção do comportamento pró-social. In: REUNIÃO ANUAL DA SOCIEDADE DE PSICOLOGIA DE RIBEIRÃO PRETO, 21.1991. *Anais...* Comportamento pró-social: questões filogenéticas e ontogenéticas. Ribeirão Preto, 1991. p.110-113.

. *Socialização na pré-escola:* o papel da professora e da organização das atividades no desenvolvimento das interações sociais entre as crianças. São Paulo, 1989. Tese (Doutorado) — Instituto de Psicologia, USP.

BRANCO, A.U., METTEL, T.P.L Comportamento pró-social: um estudo com pré-escolares. *Psicologia,* v.10, n.1. p.43-61, 1984.

--------------. A interação professora-alunos na pré-escola: uma metodologia de análise. In: REUNIÃO ANUAL DA SOCIEDADE DE PSICOLOGIA DE RIBEIRÃO PRETO, 19,1989. *Resumos...,* n. 222. Ribeirão Preto, 1989a.

. Socialização na pré-escola: fatores interativos e ambientais relacionados às interações criança-criança. In: REUNIÃO ANUAL DA SOCIEDADE DE PSICOLOGIA DE RIBEIRÃO PRETO, 19,1989. *Resumos...,* n. 231. Ribeirão Preto, 1989b.

BRANDÃO, CR. Pesquisar-participar. In: BRANDÃO, CR. (Org.). *Pesquisa participante.* São Paulo: Brasiliense. 1981.

BUSSAB, V.S.R. Comportamento pró-social: questões filogenéticas. In: REUNIÃO ANUAL DA SOCIEDADE DE PSICOLOGIA DE RIBEIRÃO PRETO, 21.1991. *Anais...* Comportamento pró-social: questões filogenéticas e ontogenéticas. Ribeirão Preto, 1991. p.197-209.

--------------. Processos de ritualização: aspectos filogenéticos e ontogenéticos. In: REUNIÃO ANUAL DA SOCIEDADE DE PSICOLOGIA DE RIBEIRÃO PRETO, 18,1998. *Resumos...,* n. 132. Ribeirão Preto, 1988.

CAMPOS DE CARVALHO, M.I. *Arranjo espacial e distribuição de crianças de 2-3 anos pela área de atividades livres em creche.* São Paulo, 1990. Tese (Doutorado) — Instituto de Psicologia, USP.

CAMPOS DE CARVALHO, M.I. Organização espacial da área de atividades livres em creches. In: REUNIÃO ANUAL DA SOCIEDADE DE PSICOLOGIA DE RIBEIRÃO PRETO, 18, 1988. *Anais...* Condições ambientais na educação do indivíduo especial. Ribeirão **Preto,** 1988. p.305-310.

CARVALHO, A.M.A. *Abordagem etològica:* desenvolvimento, socialização. São Paulo, 1978. Apostila utilizada no curso de Pós-Graduação do Departamento de Psicologia Experimental. Instituto de Psicologia, USP.

--------------. Algumas reflexões sobre o uso da categoria "interação social". In: REUNIÃO ANUAL DA SOCIEDADE DE PSICOLOGIA DE RIBEIRÃO PRETO, 18, 1988. *Anais...* Avanços recentes em Psicologia do desenvolvimento. Ribeirão Preto, 1988. p.511-515.

--------------. *Alguns dados sobre a divisão de trabalho entre obreiras de Atta sexdens rubropilosa, Forel, 1908 (Hymenoptera-Formicidae), em colônias iniciais, mantidas em laboratório.* São Paulo, 1972. Tese (Doutorado) — Faculdade de Filosofia, Letras e Ciências Humanas, USP.

--------------. Critérios para a descrição e análise de relações interindividuais: revisão crítica e análise empírica. In: REUNIÃO ANUAL DA SOCIEDADE DE PSICOLOGIA DE RIBEIRÃO **PRETO,** 21,1991. *Resumos...,* n. 202. Ribeirão Preto, 1991a.

--------------. Questões conceituais. In: REUNIÃO ANUAL DA SOCIEDADE DE PSICOLOGIA DE RIBEIRÃO PRETO, 21, 1991. *Anais...* Comportamento pró-social: questões filogenéticas e ontogenéticas. Ribeirão Preto, 1991b. p.104-106.

. Relações interpessoais: o nível de recorte. In: REUNIÃO DA SOCIEDADE BRA SILEIRA DE PSICOLOGIA, 22. 1992. *Resumos de comunicações científicas...,* n.342.2. [S.I.], 1992.

CARVALHO, A.M.A., BERALDO, K.E.A. Interação criança-criança: ressurgimento de uma área de pesquisa e suas perspectivas. *Cadernos de Pesquisa,* São Paulo, v.71, p.55-61, 1989.

CATANIA, A.C. The nature of learning. In: NEVIN, J.A., REYNOLDS, G.S. (Orgs.). *The study of behavior.* Glenview: Scott, Foresman and Co, 1973.

CHAGURI, AC, RUBIANO, M.R.B., SILVA JR., V., ROSSETTI-FERREIRA, M.C. A organização de crianças em creches com diferentes enfoques pedagógicos. In: REUNIÃO ANUAL DA SOCIEDADE DE PSICOLOGIA DE RIBEIRÃO PRETO, 21,1991. *Resumos...,* n.205. Ribeirão Preto, 1991.

COPI, I.M. *Introdução à lógica.* Trad. por Álvaro Cabral. São Paulo: Mestre Jou, 1974.

CORRAZE, J. *As comunicações não-verbais.* Trad. por Roberto C. Lacerda. Rio de Janeiro: Zahar, 1982.

COUTINHO. S.M.G.. ZANNON, C.M.L.C. Estudo exploratório de interações sociais em setor de emergência pediátrica. In: REUNIÃO ANUAL DA SOCIEDADE DE PSICOLOGIA DE RIBEIRÃO PRETO, 21, 1991. *Resumos....* n. 88. Ribeirão Preto, 1991.

CREPALDI, M.A., ZANNON, C.M.LC. Hospitalização infantil: estudo da interação equipe família em unidade pediátrica. In: REUNIÃO ANUAL DA SOCIEDADE DE PSICOLOGIA DE RIBEIRÃO PRETO, 19, 1989. *Resumos...,* n. 40. Ribeirão Preto, 1989.

CUNHA, W.H.A. Acerca de um curso pós-graduado destinado ao treino da observação científica no domínio das ciências do comportamento. *Ciência e Cultura,* São Paulo, v.27, p.846-853, 1974.

CUNHA, W.H.A. Alguns princípios de categorização, descrição e análise do comportamento. *Ciência e Cultura,* São Paulo, v.28, p.15-24, 1976.

------------- . O estudo etològico do comportamento animal. *Ciência e Cultura.* São Paulo, v.27 p.262-268, 1975.

------------- . *Exploração no mundo psicológico das formigas:* uma investigação de algumas modificações de comportamento. São Paulo, 1967. Tese (Doutorado) — Faculdade de Filosofia, Letras e Ciências Humanas, USP.

------------- . *Introdução ao desenvolvimento histórico e aos princípios básicos da etologia.* Palestra apresentada no I Encontro de Etologia. Jaboticabal, 1983.

DANNA, M.F., MATOS, M.A. *Ensinando observação:* uma introdução. São Paulo: Edicon, 1982.

DELAMONT, S., HAMILTON, D.A. A pesquisa em sala de aula: uma crítica e uma nova abordagem. In: PATTO, M.H.S. (Org.) *Introdução à Psicologia Escolar.* Sao Paulo: TA. Queiroz, 1981.

DELEVATI, N.M., LIRA, B.B.P., CÉSAR, O.P., PIRES, S.G., OTTA, E. Influência do sorriso na percepção de pessoas em função de características do sujeito e do modelo. In: REUNIÃO ANUAL DA SOCIEDADE BRASILEIRA DE PSICOLOGIA, 22,1992. *Resumos de comunicações científicas...,* n.30. [S.I.], 1992.

DESSEN, M.A.S.O, METTEL, T.P.L. Interação pais-primogênito quando da chegada de uma segunda criança na família - um estudo de caso. *Psicologia,* v.10, n.2, p.27-39, 1984.

DURAN, A.P. O estudo da interação professor-aluno no Brasil: algumas tendências, algumas questões. In: REUNIÃO ANUAL DA SOCIEDADE DE PSICOLOGIA DE RIBEIRÃO PRETO, 17,1987. *Anais...* Perspectivas para estudo da relação professor-aluno. Ribeirão Preto, 1987. p.77-85.

. *Padrões de comunicação oral e compreensão da comunicação escrita na universi dade:* estudos no Nordeste. São Paulo, 1981. Tese (Doutorado) — Instituto de Psicologia, USP.

FAGUNDES, A.J.F.M. Definição e análise da resposta de sorrir em situação de leitura de textos humorísticos. *Psicologia,* v.4, n.3, p.53-108, 1978.

------------- . *Descrição e registro de comportamento.* São Paulo: Edicon, 1981.

GIL-CENEVIVA, M.S.A. *Interação professor-aluno:* estudo descritivo das perguntas dos alunos em uma sala de aula de primeira série do Primeiro Grau. João Pessoa, 1985. Dissertação (Mestrado)— Universidade Federal da Paraíba.

GIL, M.S.C.A. *Análise funcional da interação professor-aluno:* um exercício de identificação de controles recíprocos. São Paulo, 1990. Tese (Doutorado) — Instituto de Psicologia, USP.

. Análise funcional da interação professor-aluno: In: REUNIÃO ANUAL DA SOCIE DADE DE PSICOLOGIA DE RIBEIRÃO PRETO, 21,1991. *Anais...* Ciclo de estudos em análise do comportamento IV: novos problemas e novas soluções. Ribeirão Preto, 1991. p.315-321.

HALL, R.V. *Manipulação de comportamento - parte 1:* modificação de comportamento. Trad. por Waldir Bettoi. São Paulo: EPU: EDUSP. 1973.

HINDE, R.A. *Animal behaviors* synthesis of ethology and comparative psychology. New York: McGraw-Hill, 1966.

HINDE. R.A. The bases of a science of interpersonal relationships. In: DUCK, S., GILMOUR, R. (Eds.). *Personal relationships.* Londres: Academic Press, 1981. v.1.

----------------. On the design of check-sheets. *Primates,* Aichi, v.14, p.393-406. 1973.

HUIT, S.J., HUTT, C. *Observação direta e medida do comportamento.* Trad. por Carolina M. Bori. Sao Paulo: EPU: EDUSP, 1974.

JOHNSON, S.M., BOLSTAD, OD. Methodological issues in naturaliste observation: some problems and solutions for field research. In: HAMERLYNCK, LA., HANDY, L.C., MADH, E.J. (Eds.). *Behavior change-,* methology, concepts and practice. [S.l.]: Research Press, 1973.

KERBAUY, R.R. *Autocontrôlé:* manipulação de condições do comportamento alimentar. São Paulo, 1972. Tese (Doutorado) — Faculdade de Filosofia, Letras e Ciências Humanas, USP.

LEMOS, C. de. Interactional processes in the child's construction of language. In: DEUTSCH, W. (Ed.). *The child's construction of language.* Londres: Academic Press, 1981.

LI NHARES, M.B.M. *Interação professora-aluno em uma classe de 1a. série com substituição de professoras.* São Paulo, 1988. Tese (Doutorado) — Instituto de Psicologia, USP.

LINHARES, M.B.M., MARTURANO, E.M. Um método de observação e análise das estratégias maternas de ensino. *Psicologia,* v.10, n.l, p.l 1-25, 1984a.

----------------. Conteúdos verbais maternos em diferentes situações de ensino de resolução de problemas. *Psicologia,* v.10, n.3, p.41-55, 1984b.

LOPES, M.F., BUENO, J.L.O., BARNABE. "OBS": uma ferramenta para análise seqüencial e de freqüência e duração de varias categorias comportamentais. In: REUNIÃO ANUAL DA SOCIEDADE BRASILEIRA DE PSICOLOGIA, 22, 1992. *Resumos de comunicações científicas,* n.233. [S.l.], 1992.

LYRA, M.C. Transformações e construções do observador e dos dados na análise de interações mãe-bebê. In: REUNIÃO ANUAL DA SOCIEDADE DE PSICOLOGIA DE RIBEIRÃO PRETO, 21, 1991. *Anais...* Transcrição e análise de registro em vídeo. Ribeirão Preto, 1991. p.293-297.

----------------. *Transformação e construção na interação social:* a diade mãe-bebê. São Paulo, 1988. Tese (Doutorado) — Instituto de Psicologia, USP.

LYRA, M.C., GALINDO, W.C.M., CIPRIANO, A.M. A mediação construída "face-a-face" na diade mãe-bebê: a produção vocal e o sorriso. In: REUNIÃO ANUAL DA SOCIEDADE DE PSICOLOGIA DE RIBEIRÃO PRETO,20,1990. *Resumos...,* n.16. Ribeirão Preto, 1990.

LYRA. M.O. CABRAL, EA, PANTOJA, A.P.F. A produção vocal do bebê: diferenciação das interações "mãe-objeto-bebê". In: REUNIÃO ANUAL DA SOCIEDADE DE PSICOLOGIA DE RIBEIRÃO PRETO, 21,1991. *Resumos...,* n.214. Ribeirão Preto, 1991.

MACHADO, V.L.S. *Interação verbal professor-aluno,* influência de disciplinas, de expectativas do professor, da autopercepçáo do aluno e suas relações com rendimento acadêmico de escolares de terceira série. São Paulo, 1979. Tese (Doutorado) — Instituto de Psicologia, USP.

MARTURANO, E.M. Análise de componentes da interação professor-aluno. *Cadernos de Análise do Comportamento,* v.6, p.15-24,1984.

----------------. *Estudo da interação verbal criança-mãe.* Sao Paulo, 1972. Tese (Doutorado) — Faculdade de Filosofia, Letras e Ciências Humanas, USP.

MARTURANO, E.M. Interação professor-aluno: desafios à pesquisa. In: REUNIÃO ANUAL DA SOCIEDADE DE PSICOLOGIA DE RIBEIRÃO PRETO, 22,1987. *Anais...* Perspectivas para estudo da relação professor-aluno. Ribeirão Preto, 1987. p.86-90.

— ---------. A lógica do sistema de categorias e a análise do comportamento verbal. In: REU NIÃO ANUAL DA SOCIEDADE DE PSICOLOGIA DE RIBEIRÃO PRETO, 18, 1988. *Anais...* Avanços recentes em Psicologia do desenvolvimento. Ribeirão Preto, 1988. p.517-520.

-------------. Relação entre o julgamento da professora e o comportamento dos alunos — estudo de um caso. *Psicologia,* v.12, n.3, p.47-58, 1986.

MARTURANO, EM., BERTOLDO, A.A., CAMELO, A.L.P. Estudo descritivo do intercâmbio verbal em sala de aula através da análise de contingência — uma contribuição metodológica. *Psicologia,* v.8, n.3, p.19-36, 1982.

MARTURANO, E.M., MAIMONI, E.H. Comportamento da professora frente a alunos com dificuldade de aprendizagem. In: REUNIÃO ANUAL DA SOCIEDADE DE PSICOLOGIA DE RIBEIRÃO PRETO, 19, 1989. *Resumos...,* n.218. Ribeirão Preto, 1989.

McGREW, W.O *An ethological study of children's behaviour.* New York: Academic Press, 1972.

MEJIAS, N.P. *Modificação de comportamento em situação escolar.* São Paulo, 1973. Tese (Doutorado) — Faculdade de Filosofia, Letras e Ciências Humanas, USP.

MELO, CS., BPiANCO, A.U. Adaptação à pré-escola: análise das negociações adulto-criança. In: REUNIÃO ANUAL DA SOCIEDADE DE PSICOLOGIA DE RIBEIRÃO PRETO, 22, 1992. *Resumo das comunicações científicas,* n. 136. Ribeirão Preto, 1992.

MICHAEL, J. *Characteristics of a good definition.* Washington, DC, 1975. Apostila para distribuição interna na American University.

MORAIS, G.S. (Org.). *Pesquisa e realidade no ensino de primeiro grau.* São Paulo: Cortez, 1980.

---------------. Pesquisas sobre interação professor-aluno e intervenção na escola. In: REUNIÃO ANUAL DA SOCIEDADE DE PSICOLOGIA DE RIBEIRÃO PRETO, 17, 1987. *Anais...* Perspectivas para estudo da relação professor-aluno. Ribeirão Preto, 1987. p.94-98.

OBARA, CS.. AKAMINE, C, PEDRAZZOLI NETO. M„ BONILHA, R.S.A., BORTOLETTO, A.C.C., OTTA, E. Sorriso em bebês: reação à face humana e a vários tipos de degradações deste estímulo. In: REUNIÃO ANUAL DA SOCIEDADE DE PSICOLOGIA DE RIBEIRÃO PRETO. 20. 1990. *Resumos...,* n.125. Ribeirão Preto, 1990.

OLIVEIRA, Z.M.R. *Jogo de papéis-,* uma perspectiva para análise do desenvolvimento humano. São Paulo, 1988a. Tese (Doutorado) — Instituto de Psicologia, USP.

---------------. Análise de interações criança-criança a partir da perspectiva de jogo de papéis. In: REUNIÃO ANUAL DA SOCIEDADE DE PSICOLOGIA DE RIBEIRÃO PRETO, 18,1988. *Resumos...,* n.133. Ribeirão Preto. 1988b.

---------------. Conceito de papel na análise do processo interacional. In: REUNIÃO ANUAL DA SOCIEDADE DE PSICOLOGIA DE RIBEIRÃO PRETO, 22. 1992. *Resumos de comunicações científicas,* n.342.3. Ribeirão Preto, 1992.

OTTA, E., SARRA, S. Uma análise motivacional do sorriso e do riso. In: REUNIÃO ANUAL DA SOCIEDADE DE PSICOLOGIA DE RIBEIRÃO PRETO, 18. 1988. *Resumos...,* n. 131. Ribeirão Preto, 1988.

PATTO. M.H.S. A observação antropológica na interação professor-aluno: resumo de uma proposta. In: PATTO. M.H.S. (Org.). *Introdução à Psicologia Escolar.* São Paulo: TA. Queiroz, 1981.

PEDROSA. M.I.P.C. *Interação criança-criança:* um lugar de construção do sujeito. São Paulo, 1989. Tese (Doutorado) — Instituto de Psicologia, USP.

PEDROSA. M.I.P.C, CARVALHO, A.M.A. Análise qualitativa de interação entre crianças como processo de construção mútua. In: REUNIÃO ANUAL DA SOCIEDADE DE PSICOLOGIA DE RIBEIRÃO PRETO, 21, 1991. *Anais...* Transcrição e análise de registro em vídeo. Ribeirão Preto, 1991. p.285-290.

PEDROSA, G.B. Relações entre crianças pequenas de mesma idade: algumas características da dinâmica das díades. In: REUNIÃO ANUAL DA SOCIEDADE DE PSICOLOGIA DE RIBEIRÃO PRETO, 22,1992. *Resumos de comunicações científicas,* n.146. Ribeirão Preto. 1992.

PRADO, L.E.S. Fato e significado. In: REUNIÃO ANUAL DA SOCIEDADE DE PSICOLOGIA DE RIBEIRÃO PRETO, 18, 1988. *Anais...* A questão da subjetividade em Psicologia. Ribeirão Preto, 1988. p.411-417.

RAMOS, A.T.A. *Estudo descritivo das relações contingenciais no intercâmbio verbal de criança com retardo no desenvolvimento da fala, com a professora e a mãe, em situação natural.* São Paulo, 1979. Dissertação (Mestrado)— Instituto de Psicologia, USP.

RAMOS CERQUEIRA, A.T.A. Análise de relações de contingência do repertório verbal de mães e crianças num contexto Interativo. In: REUNIÃO ANUAL DA SOCIEDADE DE PSICOLOGIA DE RIBEIRÃO PRETO, 18,1988. *Resumos....* n. 196. Ribeirão Preto. 1988a.

. Estudo descritivo do repertório verbal de mães e crianças de baixo nível sócio-econômico num contexto interativo. In: REUNIÃO ANUAL DA SOCIEDADE DE PSICO LOGIA DE RIBEIRÃO PRETO, 18, 1988. *Resumos...,* n.197. Ribeirão Preto, 1988b.

REBER, A.S. *The Penguin dictionary of Psychology.* London: Penguin, 1985.

RIBEIRO, F.L. *Um estudo sobre o comportamento da fêmea durante a fecundação da colônia em Atta sexdens rubropilosa, Porei, 1908 (Hymenoptera-Formicidae).* São Paulo, 1972. Tese (Doutorado) — Faculdade de Filosofia. Letras e Ciências Humanas, USP.

RICHARDS, M.P.M., BERNAL, J. An observational study of mother-infant interaction. In: JONES, Blurton (Ed.). *Ethological studies of child behaviour.* Cambridge: Cambridge University Press, 1972.

RODRIGUES, M.A.C. *Psicologia infantil:* a independência da criança pré-escolar. São Paulo: Edicon, 1985.

ROSSETTI-FERREIRA. M.C. *Development of a method for the study of mother-child interaction during meal-time.* Londres, 1967. Tese (Doutorado).

--------------. *Mãe e criança - separação e reencontro:* observação em situação de grupo. São Paulo: Edicon, 1986.

RUBIANO, M.R.B. Ocorrência de atividades lúdicas e compartilhadas em diferentes "settings". In: REUNIÃO ANUAL DA SOCIEDADE DE PSICOLOGIA DE RIBEIRÃO PRETO, 21, 1991. *Anais...* Transcrição e análise de registro em vídeo. Ribeirão Preto, 1991. p.281-284.

------------- . *Suportes ambientais e organização social de crianças em creches.* São Paulo, 1990. Tese (Doutorado) — Instituto de Psicologia, USP.

SATO, T. Classificação numérica do comportamento I - a construção de matrizes métricas. *Psicologia,* v.13, n.l, p.55-64, 1987.

SCHERER, K-, EKMAN, P. (Orgs.)- *Handbook of methods in nonverbal behavior research.* Cambridge: Maison des Sciences de l'Homme: Cambridge University Press, 1982.

SIDMAN, M. *Tactics of scientific research:* evaluating experimental data in Psychology. New York: Basic Books, 1960.

SILVA, A.A. *Julgamento de expressões faciais de emoções:* fidedignidade, erros mais freqüentes e treinamento. São Paulo, 1987a. Tese (Doutorado) — Instituto de Psicologia, USP.

--------------. Expressões faciais de emoções: algumas questões centrais que ainda não foram pesquisadas ou foram pesquisadas de uma forma insatisfatória. In: REUNIÃO ANUAL DA SOCIEDADE DE PSICOLOGIA DE RIBEIRÃO PRETO, 17,1987. *Anais...* Pesquisas que visam emoções. Ribeirão Preto, 1987b. p. 123-129.

SKINNER, B.F. *Ciência e comportamento humano.* Trad. por João Cláudio Todorov e Rodolpho Azzi. Brasília: Ed. UnB, 1967.

--------------. Contingências do reforço. Trad. por Rachel Moreno. In: OS PENSADORES. São Paulo: Abril Cultural, 1975. v.51.

--------------. *Verbal behavior.* Englewood Cliffs (NJ): Prentice-Hall, 1957.

SLATER, P.J.B. Describing sequences of behavior. In: BATESON, P.P.G., KLOPFER. P.H. (Eds.). *Perspectives in ethology.* Londres: Plenun Press, 1973.

SOLLITTO, N.A. *Observação da interação mãe-nenê em uma situação natural.* São Paulo, 1972. Tese (Doutorado) — PUC-SP.

STADDON, J.E.R. On the notion of cause, with applications to behaviorism. *Behaviorism,* Reno (NV), v.1, p.25-63,1973.

STELLA, E.M. *A field-descriptive and experimental study of verbal behavior in one year old children.* Londres, 1974. Tese (Doutorado).

THIOLLENT. M. Notas para o debate sobre pesquisa- ação. In: BRANDÃO, CR. (Org.). *Repensando a pesquisa participante.* São Paulo: Brasiliense, 1984.

VAN HOOF, J.A.R.A.M. Categories and sequences of behavior: methods of description and analysis. In: SHERER, K.R., EKMAN, P. (Eds.). *Handbook of methods in nonverbal behavior research.* Cambridge: Maison des Sciences de l'Homme: Cambridge University Press, 1982.

VIEIRA, T. Estudo observacional sobre modalidades de ações representativas em brinquedo de faz-de-conta. In: REUNIÃO ANUAL DA SOCIEDADE DE PSICOLOGIA DE RIBEIRÃO PRETO, 18, 1988. *Resumos...,* n.128. Ribeirão Preto. 1988a.

-------------. Brinquedo de faz-de-conta: um estudo sobre interesse de crianças em objetos. In: REUNIÃO ANUAL DA SOCIEDADE DE PSICOLOGIA DE RIBEIRÃO PRETO, 18, 1988. *Resumos...,* n.129. Ribeirão Preto, 1988b.

VIEIRA, T.A.M. *Elaboração de um catálogo de categorias de comportamento:* uma contri buição para o estudo etologico do homem. São Paulo, 1975. Dissertação (Mestrado) — Instituto de Psicologia, USP.

WEICK, KE. Systematic observation methods. In: LINDZEY, G, ARONSON, E. (Eds.). *The handbook of sodai psychology.* New York: Addison Wesley, 1968. v.2.

WITTER, G.P. *Discriminação de forma oposição:* conhecimento da contingência de reforço. São Paulo, 1969. Tese (Doutorado) — Faculdade de Filosofia, Letras e Ciências Humanas, USP.

CAPÍTULO 11

A MEDIDA DA CRIATIVIDADE

Eunice M. L Soriano de Alencar
Instituto de Psicologia
Universidade de Brasília

O interesse da Psicologia pelo estudo da criatividade e pela natureza e condições que favorecem a expressão do talento criativo é relativamente recente. Ele ocorreu a partir da década de 50, fruto de vários fatores como das contribuições de psicólogos da linha humanista, tais como Rogers (1954) e Maslow (1959). Estes apontaram para o potencial humano, para a auto-realização, abordando as condições que facilitam a expressão da criatividade e destacando a saúde mental como a fonte dos impulsos criativos. Os humanistas concebem ainda a criatividade como resultando de uma interação mutuamente benéfica entre a pessoa e o ambiente. Rogers considera, por exemplo, a autonomia humana e a resistência ao controle social excessivo como condições necessárias à atividade criativa. Por outro lado, é necessário também que a sociedade possibilite à pessoa liberdade de escolha e ação, reconhecendo e estimulando o potencial criador do indivíduo.

Um outro fator que também contribuiu para deslanchar o interesse dos psicólogos por esta área foi o discurso de Guilford, quando assumiu a presidência da Associação Americana de Psicologia, em 1950. Nesta ocasião, Guilford chamou a atenção para o descaso pela pesquisa sobre criatividade por parte dos psicólogos norte-americanos, destacando, por exemplo, que dos 121.000 títulos relacionados no *Psychological Abstracts* até aquela data apenas 186 tinham alguma relação com criatividade. Neste seu discurso, Guilford lembrou a importância social da criatividade, especialmente na busca de novas soluções para os problemas enfrentados pela humanidade, e apontou para a necessidade de um estudo sistemático de seus múltiplos aspectos (Guilford, 1950).

A partir daquele momento, diferentes componentes da criatividade, como o processo criativo, o produto criativo, a pessoa criativa e o ambiente que favorece a criatividade, passaram a ser objeto de inúmeras pesquisas, sob enfoques teóricos os mais diversos. Entre outros aspectos, têm sido objeto de investigação as habilidades cognitivas que caracterizam o pensamento criativo; a história biográfica, experiências antecedentes e

traços de personalidade de indivíduos que se destacam por sua produção criativa; as características do contexto social que favorecem a expressão da criatividade e o efeito de diferentes técnicas e programas de treinamento de criatividade e de resolução criativa de problemas.

Paralelamente aos estudos realizados sobre as diferentes dimensões da criatividade, uma enorme variedade de definições surgiu na literatura. Nove anos após o discurso de Guilford chamando a atenção dos psicólogos para a necessidade de pesquisa na área, Taylor (1959) encontrou, por exemplo, mais de cem definições distintas de criatividade, definições estas muitas vezes conflitantes e dando ênfase a aspectos diversos. Isto certamente se deve ao fato de que criatividade, como inteligência, constitui-se um construto complexo e difuso que envolve muitas dimensões e facetas.

De especial interesse tem sido também o desenvolvimento de medidas para avaliar a criatividade, tema este que será examinado no presente capítulo, onde serão abordadas as diferentes modalidades de medida e questões relativas à validade, fidedignidade e condições de aplicação dos instrumentos disponíveis.

AS DIFERENTES MODALIDADES DE MEDIDAS DE CRIATIVIDADE

Uma análise da literatura relativa aos instrumentos de medida de criatividade sugere que muitos e muitos instrumentos foram construídos a partir da década de 50 não apenas nos Estados Unidos, mas também em alguns países europeus, como Itália, Suécia e Alemanha (para uma descrição de alguns dos instrumentos publicados em países europeus, veja Urban, 1990).

Uma revisão crítica de mais de 100 distintas medidas de criatividade foram feitas em anos recentes por Hocevar e Bachelor (1989), que, após examinarem os instrumentos disponíveis, classificaram-nos em várias categorias, como as especificadas a seguir:

— testes de pensamento divergente;
— inventários de atitudes e interesses;
— inventários de personalidade;
— inventários biográficos;
— nomeação por professores, pares e supervisores;
— julgamento de produto;
— auto-registro de atividades;
— realizações criativas.

Além destas, Hocevar e Bachelor incluem ainda o estudo de pessoas eminentes, o que, no entender da autora do presente capítulo, não se

constitui um tipo de medida, uma vez que, nestes estudos, distintas modalidades de medidas, incluindo desde testes de pensamento criativo e inventários biográficos até entrevista clínica, têm sido utilizadas.

Dentre estas diversas modalidades, a que tem sido mais usada em pesquisas na área de criatividade são os testes de pensamento divergente (às vezes, denominados de pensamento criativo). Observa-se que os primeiros testes desta natureza foram desenvolvidos por Guilford durante a década de 40, quando realizava estudos na área de inteligência. Estes estudos culminaram com a construção de um modelo teórico da inteligência humana, a que denominou Estrutura do Intelecto, que inclui 120 habilidades distintas, e que foi inicialmente descrito em um artigo publicado por ele em 1959 (Guilford, 1959)' e posteriormente em livro sobre a natureza da inteligência (Guilford, 1967).

Dentre as distintas operações intelectuais propostas por Guilford e incluídas neste modelo, destaca-se o pensamento divergente. Este diz respeito à produção de inúmeras respostas alternativas para uma questão e contrasta com o pensamento convergente, onde se exige do sujeito apenas uma única resposta.

Uma das contribuições mais significativas que têm sido atribuídas a Guilford diz respeito aos vários testes propostos por ele para medir as diferentes habilidades relacionadas ao pensamento divergente e a outras habilidades que também contribuem para a criatividade. Nota-se que Guilford levantou a hipótese de que subjacente à criatividade estariam pelo menos oito habilidades primárias, várias delas relacionadas a este tipo de pensamento, como: fluencia (habilidade de gerar um grande número de idéias ou respostas para um dado problema); flexibilidade (esta implica mudanças, seja no significado, na interpretação ou no uso de algo, na estratégia de se fazer uma dada tarefa ou, ainda, na direção de pensamento); originalidade (estudada através da apresentação de respostas incomuns ou remotas), sendo o critério de raridade estatística utilizado para se determinar o grau de originalidade da resposta em uma dada população); elaboração (habilidade em acrescentar uma variedade de detalhes a uma informação, produto ou esquema); redefinição, que implica transformações, revisões ou outras modalidades de mudanças na informação; sensibilidade para problemas, que se traduz por uma habilidade em ver defeitos, deficiências em uma situação onde usualmente não se percebem problemas.

Guilford destaca também que, apesar das habilidades de pensamento divergente serem aquelas que mais diretamente estão envolvidas no pensamento criativo, todos os tipos de habilidades representadas na

* Publicado no Brasil em *A Descoberta do Talento,* de Wolfle (1971).

estrutura do Intelecto podem contribuir para a produção criativa em algum ponto do processo.

Exemplos de testes propostos por Guilford para examinar algumas habilidades do pensamento criativo são:

Fluencia Associativa: pede-se ao sujeito para escrever várias sentenças de quatro palavras, devendo cada palavra começar com as seguintes letras: K_________ u ________ y_________ i_________ (mantiveram-se aqui as originais do teste em inglês).

Usos Alternativos: pede-se ao sujeito para relacionar possíveis usos para um objeto específico, como, por exemplo, para um jornal.

Conseqüências: o sujeito deve relacionar conseqüências para um dado evento hipotético. Exemplo: o que aconteceria se as pessoas não precisassem mais dormir.

Usos Inusitados: solicita-se ao sujeito para nomear o maior número possível de usos inusitados para um objeto comum, como um tijolo.

Uma outra bateria de testes de pensamento divergente largamente utilizada em pesquisas não só nos Estados Unidos, mas também em países europeus e no Brasil, são os Testes Torrance de Pensamento Criativo (Torrance, 1966, 1974). Esta bateria incluí uma parte verbal composta de sete testes e outra figurativa composta de três testes, com duas formas distintas, e visa avaliar quatro habilidades criativas: afluência, a flexibilidade, a originalidade e a elaboração.

Um exemplo de um teste de natureza verbal da Bateria Torrance consiste na apresentação de um quadro com um desenho, solicitando-se ao examinando para escrever o maior número possível de perguntas sobre o que está acontecendo na figura. Em um outro teste verbal, pede-se ao sujeito para sugerir modificações para um brinquedo, de tal forma a torná-lo mais atraente e interessante para as crianças. Entre os testes de natureza figurativa, há o de figuras incompletas, onde se solicita ao sujeito para completar linhas apresentadas, formando figuras ou objetos distintos. Um outro exemplo de teste de natureza figurativa denomina-se Círculos. Neste, 36 círculos idênticos são apresentados, e o sujeito é instruído a fazer desenhos diferentes com cada um dos círculos.

Para se pontuar a **fluencia,** conta-se o número de respostas relevantes dadas em cada teste. A **flexibilidade** é avaliada pelo número de diferentes categorias em que as respostas podem ser classificadas. A **originalidade** é pontuada com base em um levantamento prévio da freqüência de respostas dadas a cada um dos testes, recebendo as respostas apresentadas por 2% a 4% um ponto, e aquelas dadas por menos de 2%, dois pontos — não recebem pontuação em originalidade as demais respostas (aquelas dadas por 5% ou mais da população). Já a **elaboração** é pontuada pelo número de detalhes presentes em cada resposta.

Nas instruções, é enfatizado que o sujeito deverá fazer figuras ou objetos que ninguém ainda pensou, solicitando-se ainda para que dê nomes ou títulos para cada um dos desenhos realizados.

Embora os testes de pensamento divergente sejam aqueles mais utilizados em pesquisas na área e mais discutidos pela literatura relativa à medida de criatividade, outras modalidades de instrumentos foram também desenvolvidos com vista a avaliar traços psicológicos, de personalidade, motivacionais, biográficos e de interesses, que usualmente caracterizam pessoas mais criativas. Os inventários de atitudes e interesses se enquadram nesta categoria. Um exemplo seria o de alguns instrumentos desenvolvidos por Davis e Rimm (1982) e Rimm, Davis e Bien 1982). Estes são o GIFT *(Group Inventory for Finding Creative Talent),* GIFF I e II *(Group Inventory for Finding Interests)* e o PRIDE *(Preschool Interest Description).*

Alguns exemplos de itens presentes nestes instrumentos:

> Eu tenho um bom senso de humor; eu tenho uma grande variedade de *hobbies-,* eu gosto de escrever histórias; eu gosto de tentar novas abordagens para resolver um problema, etc. Estes são respondidos em uma escala de cinco pontos, que incluem "não", "em pequena extensão", "medianamente", "acima da média" e "definitivamente".

Vários outros instrumentos voltados para a identificação de atitudes e interesses foram desenvolvidos nos Estados Unidos e têm sido também citados em pesquisas realizadas na área. O que se observa, entretanto, examinando os seus itens, é que alguns deles parecem avaliar também traços de personalidade, como autoconfiança (exemplo de um item desta natureza: eu tenho confiança em minha habilidade intelectual) e curiosidade (eu sou muito curioso).

Uma terceira categoria de medida de criatividade presente na literatura são os inventários de personalidade. Segundo Hocevar e Bachelor (1989), estes inventários têm sido sobretudo utilizados por aqueles pesquisadores que caracterizam a criatividade como um conjunto de fatores de personalidade mais do que de traços cognitivos. Uma variedade de itens são incluídos nestes instrumentos, com vista a identificar a extensão em que o sujeito apresentaria traços que se associam à criatividade, como autonomia, autoconfiança, iniciativa, perseverança, espontaneidade, sensibilidade emocional, etc.

Nota-se que também Torrance, mais conhecido por seus testes de pensamento criativo, desenvolveu um instrumento *What kind of person are you?* (Torrance, Khatena, 1970), que se enquadra na categoria de teste de personalidade. Neste instrumento, os sujeitos selecionam adjetivos para se autodescrever, em um formato de escolha forçada. Os autores consideram que os indivíduos mais criativos tendem a se descrever, por

exemplo, como mais altruístas do que corteses, como uma pessoa de iniciativa mais do que como obediente, identificando-se aqueles mais criativos a partir de suas respostas.

Inventários biográficos, incluindo uma vasta variedade de itens relativos a *hobbies,* interesses, atividades durante a infância, produção criativa nos primeiros anos, história educacional e história familiar, são também relacionados na literatura norte-americana, onde inúmeros instrumentos já foram desenvolvidos com este fim. Segundo Amabile (1983), a maioria destes instrumentos foram originalmente desenvolvidos em uma base intuitiva e refinados através detestes com amostras distintas de indivíduos considerados mais e menos criativos.

Uma quinta categoria de medida descrita por Hocevar e Bachelor (1989) é a avaliação da criatividade por professores, colegas ou supervisores. Especialmente em pesquisas realizadas no contexto educacional, tem sido freqüente solicitar a professores para informar a respeito do nível de criatividade de seus alunos. Um dos instrumentos desenvolvidos para tal fim e já utilizado em pesquisas pela autora do presente capítulo (Alencar, 1974b, 1985) foi proposta por Torrance (1966) e denomina-se *Teacher's Evaluation of Creativity Sheet.* Neste instrumento, o professor recebe instruções para relacionar os cinco alunos mais fluentes e os cinco menos fluentes de sua classe, os cinco mais flexíveis e os cinco menos flexíveis, os cinco mais originais e os cinco menos originais, bem como os cinco melhores em elaborar idéias e aqueles cinco menos capazes de realizar tal elaboração. No instrumento, é feito um esclarecimento sobre o significado de fluencia, flexibilidade, originalidade e elaboração.

Torrance desenvolveu este instrumento com base em pesquisas anteriores (Torrance, 1962,1963) onde observou que os professores eram capazes de avaliar as habilidades criativas dos estudantes, desde que recebessem descrições adequadas do comportamento criativo. Em alguns de seus estudos, os alunos considerados pelos professores como mais fluentes, flexíveis, originais e mais competentes em elaboração e aqueles nomeados como menos fluentes, flexíveis, originais e com menor competência de elaboração diferiam entre si em resultados nos Testes Torrance de Pensamento Criativo. Entretanto, tanto Alencar (1974b) como Rush, Denny e Ives (1967) observaram pouca concordância entre a indicação por parte de professores e resultados em testes de pensamento criativo.

Supervisores imediatos e parceiros têm sido também solicitados a avaliar a criatividade, especialmente em pesquisas realizadas em ambiente organizacional. Alguns estudos utilizando tal modalidade são descritos por Hocevar e Bachelor (1989), como, por exemplo, os desenvolvidos por Calvin Taylor e colaboradores, através de *checklists* elaborados por estes autores.

Uma outra maneira de se avaliar a criatividade seria através de julgamento de produto. Esta modalidade envolve o uso de juizes que, utilizando-se de critérios predeterminados, avaliam o grau de criatividade de um dado produto. Uma pesquisadora que vem utilizando e defendendo o uso de tal modalidade é Amabile (1983), que realizou uma série de estudos onde uma análise objetiva do produto era feita, buscando-se quantificar a extensão em que o mesmo era criativo. Nestes estudos, uma série de dimensões acompanhadas de uma definição descritiva de cada uma delas era dada aos juizes, tendo a autora observado em seus estudos altos índices de fidedignidade entre juizes.

Amabile considera também importante, no uso deste procedimento, que uma série de requisitos sejam obedecidos, como os especificados a seguir: 1) os juizes devem ser escolhidos dentre pessoas que tenham experiência no domínio em questão, embora o nível de experiência varie de juiz para juiz; 2) é necessário que os juizes escolhidos estejam de acordo com os critérios relativos ao que seria uma resposta criativa; 3) é necessário que os juizes façam sua avaliação independentemente um do outro; 4) os juizes devem avaliar os produtos um em relação ao outro nas dimensões consideradas e não avaliá-los considerando-se um padrão absoluto para o produto em um dado domínio. A autora considera ainda que o critério mais importante para este procedimento de avaliação é que os julgamentos sejam fidedignos.

QUESTÕES RELATIVAS À FIDEDIGNIDADE, VALIDADE E CONDIÇÕES DE APLICAÇÃO DOS INSTRUMENTOS DE MEDIDA DE CRIATIVIDADE

O interesse crescente pelo estudo da criatividade observado nesta segunda metade de século contribuiu de uma maneira significativa para o desenvolvimento da área dos testes psicológicos. Isto porque uma vasta diversidade de instrumentos visando medir as diferentes dimensões da criatividade foram construídos e vêm sendo utilizados pelos estudiosos da área, embora sejam os testes de pensamento divergente aqueles mais freqüentemente usados tanto no contexto educacional como em pesquisas de natureza psicológica e educacional (Michael, Wright, 1989). Entretanto, quando se examina o progresso da pesquisa na área de avaliação deste construto nos últimos 20 anos, comparando-se, por exemplo, a literatura mais recente (Treffinger, 1987; Brown, 1989; Hocevar, Bachelor, 1989; Michael, Wright, 1989; Cooper, 1991) com alguns artigos escritos em décadas anteriores, como o de Crockenberg (1972) e o de Petrosko (1978), constata-se facilmente que muitas dificuldades ainda persistem e muitos dos problemas fundamentais ainda não foram resolvidos.

Algumas destas dificuldades se devem ao fato de que a proliferação de pesquisas em criatividade e de instrumentos de medida não foi acom-

panhada de avanços significativos na fundamentação teórica ou na organização de um corpo sistemático de conhecimento empírico, como observa Greeno (1989). Este autor aponta, por exemplo, que a pesquisa com relação ao pensamento criativo progrediu pouco se comparada com os rápidos progressos observados no estudo das estruturas cognitivas.

Ainda com relação aos testes de criatividade, Treffinger (1987) destaca que, embora haja um número muito grande de testes que se dizem "medidas" da criatividade, estes diferem entre si em inúmeros aspectos, uma vez que espelham as crenças e preconceitos de seus organizadores com relação à natureza da criatividade. Este autor vai além e lembra que, "infelizmente, a base teórica de tais instrumentos não é freqüentemente suficiente para permitir testes sistemáticos de predições diferenciais" (id. ibid., p.105).

Inúmeros foram também os autores que examinaram os dados relativos à fidedignidade e validade dos distintos instrumentos, como Anastasi (1988), Petrosko (1978), Tannenbaum (1983), Treffinger (1987), Michael e Wright (1989) e Cooper (1991), apontando para resultados inconsistentes nas pesquisas realizadas, para as dificuldades na seleção de critérios (por exemplo, na escolha de indicadores externos de comportamentos criativos), lembrando ainda a necessidade de uma maior atenção à questão da medida por parte daqueles que pesquisam criatividade. Vários aspectos têm sido também examinados com relação à fidedignidade e validade, como os especificados a seguir. Nota-se, entretanto, através de uma análise da literatura, que são sobretudo os testes de pensamento criativo que têm sido tradicionalmente considerados, não fazendo a maioria dos autores menção às demais modalidades de medidas de criatividade, como as examinadas por Hocevar e Bachelor (1989). Estes, como destacado anteriormente, organizaram uma taxonomía das medidas disponíveis, classificando-as nas diferentes categorias descritas em páginas anteriores.

Fidedignidade

Nas várias medidas de criatividade disponíveis na literatura, sobretudo nas que dizem respeito aos testes de pensamento divergente, as três distintas formas de se estimar fidedignidade têm sido encontradas. Tanto a abordagem teste-reteste — ou seja, a correlação entre a mesma forma de um teste aplicado em diferentes momentos, como formas alternativas ou paralelas de um mesmo instrumento — como a abordagem de consistência interna têm sido empregadas para estimar a fidedignidade. Entretanto, segundo Michael e Wright (1989), dependendo da modalidade de medida de criatividade, um ou outro tipo de abordagem seria mais adequado. Segundo estes autores, no caso de medidas de pensamento divergente, por exemplo, o uso de formas alternativas seria mais adequado, ao passo que, no caso de qualquer tipo de avaliação que envolva

observação, que é o que ocorre, por exemplo, quando se utilizam juizes para avaliar a criatividade de produtos, as abordagens de consistência interna são as mais apropriadas.

Como discutido anteriormente por Fleith e Alencar (no prelo), a fidedignidade nas medidas de criatividade pode ser afetada principalmente pela heterogeneidade do domínio amostrado, ou seja, pela ausência de um universo claramente definido a respeito de criatividade, do qual os itens dos instrumentos são retirados, pela amostragem de conteúdo, pela subjetividade do avaliador e pelos fatores motivacionais (tanto do avaliador quanto do avaliado) presentes na situação de avaliação.

Algumas das dificuldades encontradas com relação à fidedignidade devem-se ao fato de que há dados sobre pessoas criativas que sugerem variações substanciais da mesma em diferentes períodos em sua produção criativa. Ademais, fatores motivacionais e mesmo do contexto podem afetar o desempenho do sujeito em um teste, e isto naturalmente vai refletir em alguns índices de fidedignidade da medida. Apesar destes aspectos, no caso do Teste Torrance de Pensamento Criativo, que é o mais utilizado em pesquisa, há diversos estudos no manual do mesmo (Torrance, 1966,1974) indicando coeficientes de fidedignidade teste-reteste acima de 0,50 para as diferentes medidas. Em um de seus estudos apresentados no manual, com uma amostra de estudantes universitários, Torrance obteve coeficientes variando de 0,68 a 0,85 para os vários testes de sua bateria em um intervalo de três meses. Os estudos revistos por Torrance indicam ainda coeficientes de fidedignidade mais altos para adultos do que para crianças e maiores nas medidas de fluencia e flexibilidade dos testes verbais. Em um estudo realizado pela autora (Alencar, (1974a), em uma amostra de 159 sujeitos de 4ª e 5ª séries, obtiveram-se coeficientes variando de 0,01 a 0,56 para 12 medidas de criatividade dos Testes Torrance de Pensamento Criativo em um intervalo de quatro meses, bem mais baixos, portanto, do que os coeficientes citados na literatura.

Validade

Vários são os autores que examinaram, no contexto da avaliação da criatividade, algumas questões relativas à validade. Esta, como se sabe, é o aspecto mais importante a ser considerado, e está ancorada, segundo Wolf (apud Michael, Wright, 1989), em três questões, a saber: a) o que o teste supõe medir; b) o que o escore derivado da aplicação de um teste significa; c) como o escore de um indivíduo em uma medida se relaciona com outros fatos observáveis relativos ao indivíduo.

Com relação à validade de critério, que inclui tanto a validade concorrente como a preditiva, o grande desafio tem sido a identificação de medidas de critério relevantes e que sejam plenamente satisfatórias. No caso da validade concorrente, esta tem sido, muitas vezes, levantada através

da comparação de resultados de testes com avaliações feitas por professores e colegas, que, como vimos anteriormente, muitas vêzes apresenta dificuldades. Por outro lado, com relação à validade preditiva, especialmente no caso da Bateria Torrance, inúmeros estudos têm apontado para correlações significativas observadas entre resultados de testes obtidos em um dado momento e o registro de atividades criativas por parte dos mesmos sujeitos cinco ou mais anos após. Este aspecto foi examinado por Treffinger para os Testes Torrance de Pensamento Criativo, tendo este autor observado correlações positivas e significativas entre resultados nos testes e critérios de realizações criativas em estudos envolvendo períodos que variaram de 9 meses a 22 anos (Treffinger, 1985).

Examinando a validade de critério de testes de criatividade, Michael e Wright (1989) lembram alguns fatores que podem afetar a interpretação de um coeficiente de validade, às vezes adversamente. Estes autores lembram, por exemplo, a necessidade de que o observador ou avaliador de comportamentos criativos se previna contra o uso que, às vezes, consciente ou inconscientemente, faz de informações obtidas através da administração prévia de testes, no momento em que estiver fazendo um julgamento do desempenho do sujeito em algumas medidas do critério. Ademais, segundo estes autores, o grau de fidedignidade, tanto do preditor mas especialmente da medida de critério, é um elemento potencialmente atenuante da magnitude do coeficiente de validade. Na maior parte das circunstâncias, a fidedignidade da medida de critério tende a ser mais baixa do que aquela associada com o teste. Caso a variável de critério ou teste não for estável e consistente, isto certamente vai afetar o grau de correlação entre as medidas.

Com relação à validade de construto, tem sido lembrado que um dos requisitos para o seu estabelecimento é a validade convergente, ou seja, a demonstração de que a medida selecionada de um dado comportamento se relacione com outras medidas do mesmo construto e com outras variáveis relativas ao critério, que tenham uma relação com o construto. Um segundo requisito se refere à validade discriminativa, demonstrando que a medida selecionada é independente de variáveis de outros construios teoricamente distintos (Michael, Wright, 1989).

Um dos aspectos que têm sido discutidos com relação à validade de construto diz respeito ao fato de que uma vez que a criatividade engloba um conjunto complexo de construios interrelacionados, um dos grandes desafios reside na operacionalização de construios teóricos claros que constituam o fundamento do comportamento criativo.

Várias evidências de validade de construto têm sido apontadas para alguns dos instrumentos disponíveis na área, especialmente para a Bateria Torrance. Entretanto, há necessidade de mais pesquisas no caso da vasta maioria de instrumentos disponíveis para que os padrões de qualidade apontados pela psicometria sejam atendidos.

Quanto à validade de conteúdo, o grande desafio para os construtores de testes é atender às exigências de que o conteúdo do instrumento efetivamente abranja uma amostra representativa do domínio em questão. Este desafio foi inclusive apontado por Torrance (1974, p.21), no manual do seu teste, quando afirma:

> "Uma vez que a pessoa pode se comportar criativamente em um número quase infinito de maneiras, na opinião do autor seria ridículo tentar desenvolver uma bateria compreensiva de testes de pensamento criativo que constituísse uma amostra de qualquer tipo de universo de habilidades de pensamento criativo. O autor não acredita que alguém possa atualmente especificar o número e amplitude de testes necessários para dar uma avaliação completa ou mesmo adequada das potencialidades de uma pessoa para o comportamento criativo."

Tal fato justifica a crítica comumente feita aos testes de criatividade de que estes avaliam apenas uma amplitude muito limitada de habilidades, sendo, portanto, inadequado considerar o desempenho em um teste qualquer da área como indicador geral de criatividade. Por outro lado, nota-se que este é um problema que temos observado com freqüência na literatura, uma vez que é comum rotular como testes de criatividade algumas medidas que são essencialmente medidas de algumas poucas habilidades cognitivas tradicionalmente relacionadas ao pensamento criativo, como fluencia, flexibilidade e originalidade. O mesmo acontece em alguns instrumentos que pretendem avaliar traços de personalidade que parecem favorecer a expressão do potencial criador. O que ocorre é que certamente distintos pesquisadores estejam estudando fenômenos bem diferentes dependendo das medidas por eles selecionadas.

No nosso entender, o pesquisador da área deve estar sempre atento para o fato de que o construto criatividade é multidimensional e muito mais complexo, não devendo confundir criatividade com o que está sendo medido por estes instrumentos. Como bem lembram Hocevar e Bachelor, habilidades relativas ao pensamento divergente, atitudes, interesses, características de personalidade podem ser melhor descritas como correlatos do comportamento criativo, não devendo jamais ser consideradas como medidas de criatividade. Consideramos este aspecto sobretudo importante, uma vez que a natureza dos itens ou tarefas incluídos nos testes de pensamento criativo, por exemplo, como nos Testes Torrance de Pensamento Criativo, não conduzem a respostas criativas de alto nível, havendo ainda a questão do limite de tempo para cada item ou tarefa (de 5 a 10 minutos, dependendo da tarefa), o que é um fator que também não pode ser deixado de lado, considerando que quantidade de respostas é algo enfatizado nas instruções.

Condições de aplicação

Com relação a este aspecto, especialmente Michael e Wright (1989) lembram a necessidade de controle de alguns fatores, como os especifica-

dos a seguir: a) tempo disponível para o examinando responder ou para o observador fazer o seu julgamento; b) nível de dificuldade das palavras e maneira de apresentar as instruções, incluindo número comparável de exemplos ilustrativos para todos os que irão responder ao instrumento; c) controle de fatores que poderiam distrair a atenção do sujeito, como barulho, iluminação inadequada, clareza das instruções e ilustrações contidas no instrumento e até mesmo a seqüência e número de testes administrados.

Com relação a este último aspecto, a autora deste capítulo tem observado, por exemplo, que no caso da Bateria Torrance, a aplicação de toda a parte verbal mostra-se bastante cansativa, não só para crianças mas até mesmo para adultos. Isto faz com que muitos sujeitos não dêem o número de respostas que poderiam, especialmente nos últimos testes.

Treffinger (1987) fez uma revisão de vários estudos sobre o efeito de condições de aplicação e clima predominantes durante a mesma no desempenho em testes de criatividade. Estes estudos mostram que os resultados nestes testes são afetados pelos procedimentos utilizados durante a administração dos instrumentos. Por esta razão, é necessário uma orientação segura e mesmo um treinamento para aqueles que vão aplicar testes, no sentido de se assegurar condições adequadas e comparáveis em todas as aplicações.

A nossa experiência tem mostrado também que um cuidado especial deve ter o pesquisador no uso de testes de criatividade, quando a aplicação é grupai e não individual. Sobretudo no caso da utilização de testes de pensamento criativo para crianças e adolescentes e para sujeitos de *status* sócio-econômico baixo, o fator motivacional afeta em muito o desempenho nos testes — e, em grupo, é difícil de se controlar esta variável. O próprio Torrance, autor do teste mais utilizado de pensamento criativo, lembra que a maneira como o seu teste é aplicado afeta os resultados. Segundo ele, até o momento em que é aplicado (final de semestre, por exemplo) exerce efeito no desempenho, tendo Torrance constatado que alunos, ao final do semestre, não saíam tão bem neste instrumento como anteriormente, mesmo após um treinamento de criatividade e com alta motivação (Torrance, 1988).

Uma análise do estado da arte em medidas de criatividade nos leva, pois, a concluir que estamos ainda em um estágio bem preliminar de domínio de conhecimento nesta área, com muitas conquistas a serem feitas para que possamos clarificar muitas das questões a respeito de medidas de criatividade, que não foram ainda respondidas ou o foram de maneira incompleta ou inadequada. Como lembra Petrosko (1978), o grande desafio de se medir criatividade reside no paradoxo de se tentar construir uma maneira padronizada de se capturar algo que foge à padronização.

REFERENCIAS BIBLIOGRÁFICAS

ALENCAR. E.M.L.S. *A study of creativity training in elementary grades in Brazilian schools.* Lafayette (IN), 1974a. Tese (Doutorado) — Purdue University.

. Avaliação da criatividade de alunos por professores. *Interamerican Journal of Psy chology,* Los Angeles (CA), v.8, n.3/4, p.219-224. 1974b.

--------------- . The identification of the creative gifted by trained and untrained elementary-school teachers. In: WORLD CONFERENCE ON GIFTED AND TALENTED CHILDREN, 6,1985. *Anais....* p.57.

AMABILE, TM. *The social psychology of creativity.* New York: Springer-Verlag, 1983.

ANASTASI.A. *Psychological testing.* New York: Macmillan. 1988.

BROWN, R.T. Creativity: What are we to measure? In: GLOVER. J.A.. RONNING, R.R., REYNOLDS. C.R. (Eds.). *Handbook of creativity.* New York: Plenum Press, 1989.

COOPER, E. A critique of six measures for assessing creativity. *Journal of Creative Behavior,* Buffalo (NY). V.25. n.3. p.194-204. 1991.

CROCKENBERG. S.B. Creativity tests: a boon or boondoggle for education? *Review of Educational Research,* v.42. n.l, p.27-45, 1972.

DAVIS, G.A., RIMM. S.G. Groups Inventory for Finding Interests (GIFFI I and II): instruments for identifying creative potential in the junior and senior high school. *Journal of Creative Behavior.* Buffalo (NY), v.16. n.l, p.50-57, 1981.

FLEITH. D.S., ALENCAR, E.M.L.S. Medidas de criatividade. *Psicologia: Teoria e Pesquisa.* Brasília, [1955]. No prelo.

GREENO, J.G. A perspective in thinking. *American Psychologist* Washington, D.C., v.44, n.2, p.134-141, 1989.

GUILFORD. J.P. Creativity. *American Psychologist,* Washington. D.C., v.5. p.444-454, 1950.

----------------. 77ie *nature of human intelligence.* New York: McGraw-Hill, 1967.

HOCEVAR, D., BACHELOR, P. A taxonomy and critique of measurements used in the study of creativity. In: GLOVER. J.A.. RONNING. R.R.. REYNOLDS. C.R. (Eds). *Handbook of creativity.* New York: Plenum Press. 1989.

MASLOW, AH. Creativity in self-actualizing people. In: ANDERSON. H.H. (Ed.). *Creativity and its cultivation.* New York: Harper, 1959.

MICHAEL, W.B., WRIGHT, C.R. Psychometric issues in the assessment of creativity. In: GLOVER, J.A.. RONNING, R.R.. REYNOLDS, C.R. (Eds.). *Handbook of creativity.* New York: Plenum Press, 1989.

PETROSKO, J.M. Measuring creativity in elementary school: the current state of the art. *Journal of Creative Behavior.* Buffalo (NY), v.12. n.2. p.109-119, 1978.

RIMM. S., DAVIS. G.A.. BIEN. Y. Identifying creativity: a characteristics approach. *Gifted Child Quarterly,* Williamsville (NY), v.26. n.4, p.165-171. 1982.

ROGERS, C.R. Toward a theory of creativity. In: VERNON, P.E. (Ed.). *Creativity.* Harmondsworth: Penguin, 1982. Original de Rogers de 1954.

RUSH, R.R., DENNY, D.A., IVES. S. Fostering creativity in the 6th grade and its effect on achievement. *Journal of Experimental Education.* Washington, D.C., v.36, p.80-86, 1967.

TANNEMBAUM, A.J. Gifted children. *Psychological and educational perspectives.* New York: Macmillan, 1983.

TAYLOR, LA. The nature of the creative process. In: SMITH, P. (Ed.). *Creativity,* an examination of the creative process. New York: Hartings House, 1959.

TORRANCE, E.P. *Guiding creative talent.* Englewood Cliffs (NJ): Prentice-Hall, 1962.

--------------- . *Education and the creative potential.* Minnesota: University of Minnesota Research Bureau, 1963.

. *Torrance tests of creative thinking:* norms- technical manual. Princeton (NJ): Person nel Press, 1966.

--------------- . *Torrance tests of creative thinking:* norms and technical manual. Bensenville (Il): Scholastic Testing Service, 1974.

--------------- . The nature of creativity as manifest in its testing. In: STERNBERG, R.J. (Ed.). *The nature of creativity,* contemporary psychological perspectives. Cambridge: Cambridge University Press, 1988.

TORRANCE, E.P., KHATENA, J. What kind of person are you? *Gifted Child Quarterly,* Williams-ville (NY), V.14, p.71-76, 1970.

TREFFINGER, D.J. Research on creativity assessment. In: ISAKSEN, S.G. (Ed.). *Frontiers of creativity research:* beyond the basics. Buffalo (NY): Bearly. 1987.

URBAN, K.K. Recent trends in creativity research and theory. *Western Europe,* v.1.n.O, p.99-113,1990.

WOLFLE, D. *A descoberta do talento.* Rio de Janeiro: Lidador, 1971.

CAPÍTULO 12

ESTRATEGIAS E MEDIDAS EM ANÁLISE DE CONTEÚDO

Edson A. de Souza Filho
Universidade Estadual do Norte Fluminense, Campos, RJ

Podemos dizer que várias modalidades de análise de conteúdo vêm sendo utilizadas por pesquisadores e profissionais de mais de uma especialidade em ciências humanas, entre os quais, psicólogos e sociólogos. Ou seja, todas as vezes que dados simbólicos, verbais ou não, são utilizados para estabelecer inferência sistemática, seja a respeito de um processo mental, seja sobre uma contingência comportamental, estamos diante de uma análise de conteúdo. Contudo, desde seu aparecimento acadêmico nos anos 20 (Bardin, 1977), diferentes disciplinas e respectivas correntes teóricas têm preferido desenvolver, de acordo com necessidades intelectuais ou culturais, suas próprias definições e técnicas correspondentes. No âmbito deste capítulo, nossa intenção é apresentar algumas das definições, posturas e técnicas mais empregadas neste campo, sobretudo, por psicólogos e psicossociólogos.

Embora alguns psicólogos eminentes tenham reconhecido esta técnica, o fato de que a experimentação e a quantificação tenham passado a ocupar um lugar privilegiado em termos de estratégias e critérios gerais de pesquisa acabou diminuindo sua presença em alguns manuais e revistas mais respeitados a partir dos anos 70. Contudo, defendemos a posição de que a análise de conteúdo (A.C.) constitui-se necessidade permanente, tanto como facilitadora de inovação em ciências humanas quanto como complementação da administração da prova.

Assim, para Cartwright (1974), "uma das habilidades básicas exigidas de um psicólogo social é analisar material simbólico ou 'qualitativo'. Parte considerável da pesquisa sociopsicológica consiste em classificar e interpretar produtos verbais e outros símbolos de indivíduos ou grupo de pessoas". Ou seja, ainda que a construção de uma teoria psicológica e/ou psicossocial não possa ser suficientemente apoiada em procedimentos de tipo classificatòrio, necessariamente ela deve ter seus enunciados baseados em boas descrições de elementos de realidade simbólica, a serem organizados em tipologias e princípios de ordem mais geral, cada

vez mais sistemáticos, antes de qualquer veleidade experimental. Contudo, outro aspecto levantado por Cartwright parece-nos relevante. Refere-se ao fato de que os fenômenos tratados por disciplinas de ciencias humanas são mediados ou, mesmo, estreitamente vinculados às dimensões simbólica e verbal, tornando necessário um adequado trabalho dos mesmos em instrumentos e procedimentos adotados. Um certo reconhecimento do problema pode ser observado através da crescente valorização do trabalho intercultural, obrigando pesquisadores a desenvolver estratégias de validação em que o qualitativo ou culturalmente particular só é superado por meio de um ir e vir incessante entre conceitos descritivos ou operatórios e conceitos explicativos mais gerais. No quadro de países onde falta uma experiência científica mais consolidada em ciências humanas, o perigo de adoção de modelos construídos a partir de problemáticas e realidades estranhas é ainda maior, pois eles começam sua formação a partir de modelos teóricos oriundos de realidades não diretamente acessíveis, apesar da riqueza de detalhes de muitos relatórios. De todo jeito, em termos empíricos, a universalidade de teorias e modelos só poderá ser contestada a partir de uma revisão de fatos novos que são construídos, via de regra, através de um contato do pesquisador ou profissional com uma realidade particular, a qual, após ser filtrada e reconstituída em seus elementos, torna possível o novo conhecimento. Sabemos que os pesquisadores relutam em apresentar relatórios em que métodos considerados menos rigorosos foram usados, prejudicando a aprendizagem de produção de conhecimento novo.

Mais recentemente vêm sendo levantadas questões relativas ao método de coleta de dados, no sentido de impor problemáticas de pesquisa construídas pelo pesquisador através de instrumentos usados, além de estilo e postura adotados. Assim, uma psicóloga social italiana que havia realizado mais de um levantamento com escalas de atitudes a respeito do doente mental na Itália, constatando uma crescente favorabilidade atitudinal entre faixas etárias à medida em que aumentava a idade, surpreendeu-se ao verificar que os mesmos adultos que eram favoráveis ao doente mental já não o eram através de desenhos livres sobre o mesmo tipo de pessoa (De Rosa, 1987). De modo que não basta que os instrumentos sejam validados em termos semânticos, mas saber se eles são capazes de mapear a totalidade do campo de subjetividade do sujeito que se pretende observar — o que exige uma forma de abordagem inicial mais aberta, bem como a necessidade de A.C. antes da elaboração de instrumentos mais estruturados.

Entretanto, é preciso lembrar que as principais técnicas e procedimentos em A.C. foram criados à sombra de alguma teoria psicológica e/ou psicossocial preexistente. Nesse sentido, cada uma das grandes correntes intelectuais psicológicas e sociais modernas contribuiu de algum modo para o desenvolvimento de A.C, assim como acabou esta-

belecendo limites que foram sendo enfrentados paulatinamente por outras correntes, resultando em um campo aparentemente heteróclito, mas bastante fértil de instrumentos de trabalho. Assim, tivemos analistas de conteúdo de orientação comportamental cujas contribuições maiores foram no sentido de descrever os aspectos manifestos de verbalização, oral ou escrita, visando, em grande parte, à predição de comportamentos em relação a realidades ambientais específicas. Poderíamos lembrar os trabalhos sobre comunicação de mensagem política, feitos por Lasswell e colaboradores (1949), culminando com os de Berelson nos anos 50 (1952, 1954), que apresentam uma síntese da técnica em várias aplicações. Tais trabalhos adotaram, via de regra, uma concepção segundo a qual o texto deve ser recortado, em sua totalidade, em moléculas relacionadas a contingências reais ou simbólicas. Os aspectos cognitivos, discursivos ou epistemológicos do sujeito sendo, obviamente, deixados de lado, de acordo com a posição comportamental daquela época que preconizava ênfase menor no esclarecimento da complexidade da função simbólica, tanto individual quanto social. Os resultados obtidos a partir desse procedimento foram relatórios que descrevem com alto nível de objetividade repertórios comportamentais, sobre os quais nos deteremos mais adiante.

Posteriormente, psicólogos inspirados pelas correntes gestaltistas, desenvolvidas em torno da noção de consistência cognitiva, como Osgood, se esforçaram por criar e aperfeiçoar técnicas que, em última análise, permitissem fazer inferências de processos mentais individuais. A concepção dessa corrente é a de que possuímos estruturas e elementos de cognição cuja análise só pode ser feita em conjuntos de partes inter-relacionadas. A análise de contingências do autor mencionado é uma das mais conhecidas no campo. Mas outros autores, seguindo orientação que, segundo Piaget, poderia ser considerada estruturalista, desenvolveram, na mesma época, procedimentos de A.C. similares, ainda que com objetivos diferentes. Tratava-se de inventariar conjuntos de símbolos e significados, significantes e suportes materiais (re)produzidos por sujeitos em diferentes circunstâncias. Tal procedimento procurou trabalhar de modo objetivo, incorporando, portanto, o rigor descritivo realizado pelo campo comportamental, acrescentando um interesse por aspectos de significação do conteúdo de comunicação, sobretudo da cognição e/ou estruturas culturais, suas metas gerais mais importantes. Neste último aspecto, sabemos que houve tendência de esse grupo considerar o fenômeno do conhecimento enquanto tratamento de informação que, se implicou abertura maior comparativamente às correntes comportamentais e gestaltistas iniciais, tendeu, igualmente, de um lado, a esvaziar o aspecto social do mesmo e, de outro, a ignorar os aspectos criativos ou considerados mais complexos da função simbólica.

Um terceiro momento da A.C. poderia ser caracterizado por uma maior preocupação com os aspectos discursivos e criativos da produção

de conteúdos, o que fez com que seus procedimentos de trabalhos adotados se caracterizassem por uma maior abertura, no sentido de considerar o texto de modo mais completo, incluindo aí análise de outras dimensões da produção de linguagem, psicológicas, psicossociais e/ou sociais. Tal concepção de A.C. tendeu a realizar trabalhos sem partir de sistemas simbólicos apriorísticos, preferindo observar o fenômeno como se apresenta em situações sociais naturais. Como exemplo, poderíamos lembrar a técnica de entrevista não-diretiva que, ao fazer uma crítica à tendência tradicional de algumas abordagens clínicas no sentido de interpretar a fala do entrevistado segundo um modelo preexistente, propunha um modo mais aberto de tratar o conteúdo de comunicação, particularmente freqüente entre profissionais que adotam modelos teóricos da personalidade como a Psicanálise. Tratava-se de fortalecer a autonomia da pessoa a partir de integração de dimensões da personalidade desconhecidas ou evitadas, sistematicamente, como vivências de afetos e significados organizados pela própria pessoa, o terapeuta devendo adotar uma postura mais de "espelho" do que intérprete da realidade pessoal do paciente segundo tal ou qual posição. Procedimento semelhante foi adotado pelo uso de videoteipes no trabalho de Wicklund e Duval (1972), com a técnica de aumento da auto-objetividade de si, na qual o sujeito se observa através de videoteipes e de análise oferecida por psicoterapeuta. Em Psicología Social, a abordagem de representações sociais adota A.C. aberta para observar o conhecimento informal em relação a objetos específicos. Neste último caso, o analista não só procura inferir elementos de conteúdo isolados e combinados, como se faria em qualquer análise de comportamento e estrutural, mas, igualmente, considera-os como expressão da autonomia cognoscitiva do sujeito, que também procura transformar a realidade social e natural. Para tanto, foi necessário um tratamento de aspectos formais do discurso como seqüência da fala, seu grau de concretude/abstração, assim como sua inserção em dinâmicas históricas mais amplas, etc.

Os três momentos principais de A.C, brevemente lembrados, caracterizam-se por procedimentos que, curiosamente, foram sendo incorporados de forma gradativa, à medida que surgiram historicamente, conforme pretendemos tratar na próxima seção.

CONCEPÇÕES E MÉTODOS EM A.C.

Conforme dissemos, houve um contínuo aproveitamento dos avanços realizados em cada uma das técnicas e concepções metodológicas em A.C. Contudo, tal afirmação não implica dizer que os três tipos de A.C. mencionados foram sendo abandonados no transcorrer da história, uma vez que eles continuaram e continuam sendo empregados por profissionais e cientistas com relativa autonomia, conforme exporemos a seguir.

Em termos de concepção, seria muito leviano considerar o primeiro momento de A.C. como apenas comportamental, mas devemos admitir que foram seguidores de tal sistema teórico-metodológico que levaram às últimas conseqüências, sobretudo em termos metodológicos, as principais contribuições realizadas no que hoje se convencionou chamar de análise categorial (Bardin, 1977). De todo jeito, não é nossa intenção no contexto deste capítulo apresentar ou discutir as relações existentes entre os tipos de técnica de A.C. e seu parentesco com tal ou qual modelo teórico específico, o que seria uma tarefa mais apropriada para um historiador do campo. Nosso objetivo é mais prático, e visa, basicamente, traçar a especificidade da técnica em face das demais em termos de "inspiração" teórica, acompanhadas por exemplos de trabalhos de diferentes autores.

Entre as principais contribuições para a A.C. categorial poderíamos identificar, de modo geral, uma concepção de que o dado simbólico é importante para a produção de um certo conjunto de fenômenos, individuais ou sociais, o qual é usado como elemento descritivo do conhecimento objetivo do profissional ou pesquisador. Ou seja, a redução categorial nesse tipo de procedimento não quer saber das implicações do simbólico enquanto racionalidade ou interação social, mas, unicamente, incluí-las em sua atividade de conhecimento como um epifenômeno cuja realidade se inscreve em funções ou determinações de outra ordem. Nesse caso, existem duas estratégias gerais de trabalho. A primeira consiste em observar o simbólico a partir de um sistema categorial previamente montado pelo analista, podendo ser derivado de um sistema teórico que se pretende validar ou operacionalizar a respeito de um fenômeno a ser descrito a partir de observação de incidência freqüencial de conteúdos em diferentes sujeitos e condições, enquanto a segunda, por sua vez, caracteriza-se pela busca de categorias a partir da observação da realidade, constituindo-se, portanto, exploração de um determinado conjunto de fenômenos para a construção de um novo conhecimento.

Em termos de constituição de dados, ambas as estratégias mencionadas costumam sistematizar a observação a partir de procedimentos tais como uso de questionários estandardizados ou de material simbólico produzidos em condições similares ou diferenciadas. Assim, são muito variadas as possibilidades de utilização desta técnica, dependendo muito da inventividade do pesquisador em saber adaptá-la às circunstâncias da realidade que pretende circunscrever e organizar. No caso de análise categorial aprioristica, temos alguns exemplos clássicos em ciências humanas. Durkheim (1951), em seu estudo sobre as causas sociais do suicídio, usou os atestados de óbito de pessoas que haviam se suicidado na França durante um certo período da história. Como os atestados de óbito naquele país costumavam apresentar um pequeno histórico biográfico do sujeito, ele pôde apoiar sua hipótese de que havia crescimento de taxas de suicídio nos momentos de desintegração social, como crise

política, divórcio, desemprego, o que designou como um estado social de anomia. Outro estudo conhecido é o de McClelland (1953), interessado em observar o papel do motivo para realização entre culturas. Ele analisou estórias contadas para crianças de povos que conseguiram atingir um certo grau de realização e pôde sustentar sua hipótese. No laboratório, Moscovici (1967) pretendia comparar diferentes condições de produção e comunicação simbólica em termos de situação espacial, tais como situação face a face, lado a lado, um de costas para o outro e separado fisicamente, mas não oralmente. Ele pôde constatar, por meio de A.C., que, formalmente, o vocabulário e outras características da linguagem eram modificados pelas referidas condições.

Nos três estudos acima lembrados, os diferentes autores haviam partido de hipóteses básicas, as quais, por sua vez, haviam fornecido categorias simbólicas descritivas para analisar o material constituído, gerando freqüência a respeito dos conteúdos que enfocavam. No caso do estudo de Durkheim mencionado, o conteúdo foi a presença e ausência de eventos comportamentais e mentais, que eram categorizados segundo os tipos de suicídio egoistico, altruistico ou anêmico. A partir de estudo estatístico e demográfico correlacionai, ele pôde avaliar a referida hipótese. Note-se que o dado simbólico aqui se reduziu à informação de um evento vivido pelo sujeito observado através de sua certidão, sem constituir um estudo do simbólico em si, mental ou interativo, como o próprio Durkheim propôs mais tarde em seu trabalho sobre as formas elementares da vida religiosa. Já no estudo de McClelland, temos o uso de A.C. em que o próprio material simbólico tem papel de variável independente inferida, no caso de estórias e inscrições que supôs terem produzido motivo para a realização em indivíduos e grupos pertencentes atais culturas. Por último, o trabalho de Moscovici, um caso raro de utilização de A.C. em delineamento experimental, temos o simbólico analisado enquanto tipos de respostas em seu aspecto expressivo e formal em relação a diferentes variáveis ou condições físico-espaciais no processo de comunicação.

Conforme dissemos acima, as técnicas de A.C. da primeira fase poderiam ser caracterizadas pelo uso limitado do texto e por uma menor preocupação em sua inserção em termos mentais e sociais mais amplos. Eventualmente, como é o caso de algumas análises feitas durante a Segunda Guerra Mundial, o sujeito produtor do texto podia ser um país inteiro, mas era tratado como um indivíduo, sobre o qual interessava fazer uma inferência para descobrir suas intenções ou probabilidade de uma ação futura. Entre os vários autores que mais contribuíram a partir dessa concepção, destacamos Lazarsfeld(1948), Lasswell (1949), Berelson, (1952, 1954) e Katz (1960). As maiores contribuições deste grupo de autores foram no sentido de lançar as bases de uma técnica em termos de procedimento objetivo e sistemático, o que foi adotado pela maioria dos principais especialistas que surgiram posteriormente (Holsti, 1969; Bardin, 1977/1992;

Mucchielli, 1979; Krippendorff, 1980; Weber, 1980; Ghiglione, Blanchet, 1991). De todo jeito, podemos dizer que a ênfase na objetividade segundo esta concepção de A.C. fez com que a análise se limitasse ao aspecto manifesto e incontroverso da realidade simbólica, tal como palavras, deixando de lado, por exemplo, seu aspecto conotativo, singular, entre outros. Os trabalhos mais conhecidos dessa concepção também se notabilizaram por preocupação em estabelecer inferência a partir do emprego de freqüências. Ou seja, a idéia subjacente é que, quanto mais freqüente ou presente um dado significado em material analisado, maior sua relação com uma certa realidade psicológica e/ou social.

Com o advento das teorias de consistência cognitiva e o estruturalismo, preocupadas não mais em investigar as partes do fenômeno cognitivo e social, mas em suas relações e processos, surgiu outra concepção e método de A.C.. Agora, tratava-se de levar em consideração a combinação entre os elementos simbólicos contidos na linguagem para observar invariantes associados ou dissociados por indivíduos particulares ou uma comunidade cultural e lingüística. Um trabalho desta tendência de A.C. é o estudo de Osgood (1959) sobre a Análise de Avaliação de Asserção, cujo objetivo principal é inferir atitudes a partir de textos, incluindo o aspecto avaliativo das significações em termos de direção e intensidade. Assim, uma vez reunidos os **objetos de atitudes** presentes no texto, são inventariados os **termos avaliadores comuns,** considerados como relativamente estáveis e partilhados na cultura, os quais são ligados aos objetos de atitudes por **conectores verbais,** que indicam a intensidade da conexão entre objetos e avaliação. Poderíamos tomar como exemplo a seguinte frase obtida em pesquisa que realizamos entre moradores da cidade de Brasília (Souza Filho, 1990/1991):

"Eu gosto da cidade de Brasília, mas acho que seus moradores são frios e pouco amistosos."

Objetos de atitude: cidade de Brasília, moradores.
Avaliadores comuns: gostar, frios, amistosos.
Conectores verbais: muito, são, pouco.

Osgood partiu do pressuposto de que a maioria dos avaliadores e conectores são facilmente identificáveis por serem regulares e comuns numa dada cultura, cabendo ao analista levantar tais elementos para conhecer seus invariantes subjetivos, suas estruturas simbólicas. De modo que linguagem e atitudes são aqui consideradas, sobretudo, em seu aspecto motivacional e coletivo, sem levar em conta sua função social particular enquanto ação visando metas. Destarte, os avaliadores são codificados enquanto direção e grau de associação/dissociação com os objetivos, em termos de favorabilidade/desfavorabilídade e respectiva intensidade, classificável em escala bipolar que varia de +3 a -3. Assim, no exemplo em foco, "gostar" e "amistosos" são avaliadores favoráveis, e

"frios", desfavorável; "muito" e "são" são conectores de intensidade máxima, e "pouco", intermediária.

Ou seja, os estudos das relações entre conteúdos não estão apenas interessados na freqüência de significações expressa através de temas, mas na presença/ausência simultânea entre os mesmos. Para tanto, a fragmentação das unidades de registro será acompanhada por recortes de unidades de contexto maiores, no interior das quais observam-se as relações em cada conjunto de palavras, tal como Osgood (1959) propôs em sua análise de co-ocorrência quando o texto é contínuo (fala de um sujeito) ou um documento separado (exemplar de um jornal diário, uma entrevista numa série delas), etc, para, em seguida, registrar-se em uma matriz de contingência as medidas significativas estatisticamente em termos de associação e dissociação, as quais são interpretadas como representando relações mentais. Ainda que se possa utilizar tal técnica com outros propósitos intelectuais, fica a dúvida quanto à natureza do fenômeno, se é uma associação meramente formal de um léxico mais prenhe ou mecanismo inconsciente individual, cabendo ao pesquisador bem informado, portanto, isolar as explicações que possam competir com a que ele pretende demonstrar uma correlação ou nexo causal mais forte.

Caberia ainda mencionar que, na obra de Osgood, voltada para o papel do significado em termos psicológicos, duas preocupações básicas se manifestaram: uma, de desenvolver instrumentos relativamente flexíveis para abarcar um campo semântico mais amplo, como a Análise de Avaliação de Atitudes a partir de textos ou a técnica de Análise de Contingência entre significados, brevemente apresentadas, e outra, mais fechada, cujo propósito é buscar universais semânticos para a vivência subjetiva individual em diferentes culturas (Osgood et al., 1957).

No âmbito da semiologia e da Psicossociologia do conhecimento, vale destacar o trabalho de Barthes, que, embora inicialmente preocupado com o fenômeno simbólico enquanto produção de história (1972), dedicou-se longamente ao estudo de estruturas transistóricas, como seu livro sobre o sistema da moda (1967). Até certo ponto, algumas preocupações contidas neste modelo de A.C. reaparecerão posteriormente, tais como um maior cuidado com as complexidades internas da vida simbólica, bem como o papel da singularidade discursiva.

A partir dos anos 60, começam a surgir algumas tendências de A.C. que se caracterizam, de um lado, por uma maior sistemática, no sentido de buscar recortar o texto incluindo o contexto como condição de produção simbólica e, de outro lado, por uma maior abertura em relação ao repertório, encarando-o não apenas em sua dimensão funcional ou estrutural, mas enquanto expressão de uma subjetividade particular, situada social e historicamente. Tal será a postura adotada por Moscovici (1961/1976) em seu trabalho sobre representações sociais, e por Fishbein e Ajzen (1975),

sobre atitudes, crenças e comportamentos, apesar de algumas diferenças de ênfase entre ambas as teorias.

Como se sabe, os trabalhos acima referem-se a contribuições realizadas segundo a perspectiva psicossocial. Considera-se, hoje, que, apesar de existir enorme acúmulo de informações sobre a vida simbólica oriundas da Psicologia e da Sociologia e, particularmente, da Sociologia do Conhecimento, tais abordagens teórico-metodológicas são insuficientes, isoladamente, para a compreensão de fenômenos em questão (Souza Filho, 1990). Daí a necessidade de novos modos de trabalho, articulando em modelos teóricos mais de uma dimensão de condição de produção simbólica. No caso do modelo de Moscovici, a preocupação básica é saber como diferentes grupos elaboram o conhecimento social informal. Hã, portanto, dois eixos principais de trabalho: de um lado, trata-se de considerar os critérios sociais e/ou grupais usados e, de outro lado, em conexão com o primeiro eixo, os aspectos plásticos ou objetais empregados. Por critérios sociais, entende-se normas, ideologias, crenças e mitos de ordem mais conceituais, que serviram de critérios de orientação para pensar e simbolizar, e, por aspectos plásticos, de ordem mais imagística e concreta, aquelas dimensões do objeto selecionadas bem como a organização ou forma dada ao mesmo, podendo ser espacial, hierárquica ou outra. O produto simbólico final, a representação social, seria a combinação dos dois eixos mencionados, podendo variar tanto quantos grupos existem sobre a face da Terra. Daí a necessidade de A.C. para a observação de tal fenômeno, podendo, obviamente, ser seguida por levantamentos que operacionalizem tais variáveis simbólicas e plásticas em variáveis numéricas (Doise, 1992).

Na mesma tendência de combinar as dimensões sociais e psicológicas, agora para entender as atitudes, crenças e comportamentos individuais, está o modelo de Fishbein e Ajzen, exposto em outra parte deste livro. Para prever o comportamento de um indivíduo ou grupo, os autores em foco propõem circunscrever, entre outros, 1) o objeto de atitudes e crenças, 2) os personagens, reais ou imaginários, que estão influenciando a manutenção de crenças e atitudes e 3) a intenção de comportamento em relação ao objeto. Tais objetivos farão com que, para cada estudo de atitude e crença, se faça um instrumento específico, exigindo, entre outras coisas, o que se denomina levantamento de atitudes e crenças, que nada mais é que uma A.C. combinada às dimensões sociais, ainda que esses autores não se refiram à técnica usada nestes termos.

Outras concepções de A.C. foram elaboradas, notadamente aquelas correntes que se denominam como análise de discurso. Entretanto, não caberia no espaço deste capítulo apresentá-las de modo mais exaustivo. Uma peculiaridade comum a esses trabalhos, que, até certo ponto, difere da A.C. mais corrente, é a intenção de levar em conta muitas dimensões

do fenòmeno simbólico, forçando a operar com unidades de análise maiores que a frase, às vezes prejudicando a operacionalização e quantificação de dados. Mesmo assim, Potter e Wetherell (1987) conseguiram incluir contexto social e cognitivo em seus estudos de discurso com bastante acurácia, baseando-se em conhecimentos oriundos da Psicologia e Etnometodologia. Outras abordagens, ainda, dão grande ênfase à Lingüística, Sociologia ou História, tais como as de Pécheux (1966) e Maingueneau (1987), entre outros.

DIFERENTES ETAPAS DEA.C.

Em relação à estratégia de observação que pretende analisar o material simbólico a partir do dado bruto ou análise exploratória, ela poderia ser organizada, segundo Bardin (1977/1992), em seis etapas sucessivas: pré-análise, codificação, categorização, tratamento de dados, inferência e interpretação.

A pré-análise refere-se àquela fase em que o analista, diante de uma certa problemática psicológica ou psicossocial, começa a intuir quais seriam os possíveis parâmetros e/ou variáveis importantes para a existência da mesma, passíveis de serem observados a partir do dado simbólico manifesto. Trata-se de operacionalizar, de modo sistemático, um procedimento de análise. Para tanto, é preciso partir de algumas evidências preliminares que facilitarão a escolha do material a ser usado e a formulação de hipóteses ou objetivos de análise. Como ilustração, poderíamos pensar na situação de um profissional de Psicologia que se depara com a "coincidência" de que pessoas que têm dificuldade de expressar hostilidade ou crítica em relação a figuras que representam autoridade ou posição superior apresentam, igualmente, tendência de manifestar agressão em relação a minorias. Uma possibilidade de análise seria formar dois grupos com níveis diferentes de hostilidade em relação a minorias e submetê-los a uma mesma lista de perguntas relativas a todas as suas possíveis experiências subjetivas com autoridades durante a infância. O passo seguinte seria encontrar formas ou indícios de experiências infantis de ausência de crítica e bloqueio de expressão de hostilidade em relação a autoridades, assim como de agressão em relação a sujeitos minoritários. Um questionário com perguntas abertas poderia ser preparado de modo a obter tais índices, como foi feito no estudo coordenado por Adorno (1950).

Em estudo que realizamos sobre a revista *Sabrina* (Souza Filho, no prelo), partimos da idéia de que ela exercia, principalmente para leitores femininos, função de realização, no plano da fantasia, de desejos sexuais. Para observarmos a revista de modo sistemático, fizemos uma amostragem a partir do conjunto de publicações existentes até então e começamos a analisar todas as descrições relativas aos personagens principais, em

torno dos quais girava a trama das estórias. Em seguida, solicitamos a grupos de sujeitos de sexo feminino e masculino que lessem as mesmas publicações previamente analisadas e pedimos que relatassem o conteúdo da leitura, confirmando nossas expectativas.

Assim, uma vez definido o universo sobre o qual se vai trabalhar, precisamos constituir uma amostra do material a ser analisado. Algumas regras básicas para a obtenção do material são:

1) **Exaustividade,** segundo a qual todas as partes do mesmo devem ser incluídas para evitar seletividade.

2) **Representatividade,** ou seja, a regra de levar em conta a distribuição de todas as respectivas partes constituidoras do universo em relação ao qual se pretende generalizar. No exemplo citado, pretendíamos saber se a revista tinha apresentado alguma modificação no tratamento de assuntos analisados ao longo do tempo, necessitando, portanto, de se fazer um sorteio de publicações entre cada período histórico de sua existência, o que permitiria, além disso, diminuir a quantidade de material a ser analisado, sem prejudicar a exaustividade e representatividade pretendidas.

3) **Homogeneidade,** a exigência de que a escolha do material é definida explicitamente, sem incluir heterogeneidades que não sejam controláveis por meio de amostragem, tanto em termos de condição de produção quanto do tipo de material ou população.

4) **Pertinência,** refere-se à obrigatoriedade de que o material sobre o qual se trabalha seja adequado para a verificação da hipótese ou realização do objetivo. Em última análise, o grau de adequabilidade do material é dado pela possibilidade de que ele ofereça boa conjugação entre a "leitura" do mesmo por parte do analista e o conteúdo observável, revelando algo até então imprevisto ou facilitando a obtenção dos objetivos pretendidos. Alguns acreditam mais na capacidade heurística da própria teoria ou conceito, outros, mais no dado novo. No caso de A.C. exploratória, há amplas possibilidades de descobertas, pois se descreve de modo ordenado a partir de criação de categorias inicialmente "descritivas", podendo tornar-se, posteriormente, elementos para a construção de um novo conhecimento, de modo que o trabalho de escolha do material a ser analisado muitas vezes acompanha a codificação e categorização.

Em termos de codificação, a tarefa poderia ser resumida na escolha do tipo de recorte ou unidade de registro do dado simbólico, que inclui a definição de categorias e a quantificação dos mesmos, se for o caso.

Os recortes mais usados em A.C. são o **tema,** em geral afirmações sobre um assunto feitas por meio de palavras, frases ou períodos inteiros, os quais aparecem em um texto através de idéias nucleares. Psicólogos

e psicossociólogos trabalham com o tema para fazer inferências a respeito de atitudes, crenças, representações sociais, valores, ideologias, estados emocionais, entre outros.

A **palavra** em seu aspecto formal, tais como verbo, substantivo, advérbio ou adjetivo, pode ser outra unidade de registro usada para inferir estados mentais ou epistemologia. O estudo de Osgood e Walker (1959) sobre o estilo expressivo de cartas de suicidas mostrou que características formais, como o grau de estereotipia (repetição, pobreza lexical), desorganização (erros, rupturas) e conflito (formas verbais complexas, ambivalência), estavam relacionadas ao estado motivacional do sujeito. Eles compararam cartas de indivíduos suicidas que realmente haviam tentado se suicidar com as escritas por um grupo de sujeitos solicitados a imaginar uma carta de despedida antes de se suicidarem.

O **objeto** ou **referente** pode ser, de modo geral, qualquer entidade, humana ou não, imaginária ou real, sobre a qual é construído um texto a ser analisado. Nesse sentido, no caso do estudo sobre atitudes, representações sociais, entre outros, é importante recortar o texto em torno desses elementos.

O próprio **documento** pode se tornar uma unidade de registro quando ele é tomado globalmente em relação a certa característica simbólica. O estudo de Chombart de Lauwe (1979) sobre imagens da infância empregou tal procedimento.

Além de unidades de registro, alguns analistas situam os conteúdos em relação a unidades de contexto para seu melhor esclarecimento. Assim, o contexto aqui pode ser qualquer aspecto do texto ou extra-texto que participa indiretamente do processo ou fenômeno que se pretende inferir. No caso de linguagem oral como a obtida em entrevista aberta, as expressões faciais e corporais são cruciais para a total descodificação de mensagem, conforme mostraram os trabalhos da escola de Palo Alto sobre a comunicação paradoxal. Em termos de texto, o trabalho de Osgood sobre contingência entre conteúdos, já mencionado, usou como unidade de contexto partes do texto total, tais como cada certo número de palavras, dentro de cujos limites as regularidades simbólicas são analisadas. No estudo que realizamos sobre personagens de revista de romance, separamos situações onde cada personagem aparecia sozinho ou acompanhado do parceiro.

Entre as formas de quantificação adotadas em A.C., as mais comuns são: **presença** (ou ausência) de um conteúdo em conjuntos ou partes de um material e **freqüência,** que pode ser em termos absolutos ou ponderados se o analista considera que os conteúdos não têm o mesmo valor. Assim, como já dissemos, o raciocínio que sustenta o uso de freqüência é o de que quanto mais regular ou maior a quantidade de presença, maior

sua relação com determinada população ou condição de produção. Contudo, suponhamos que alguma pessoa seja mais prolixa que outra, explicitando em maior quantidade um certo símbolo ou conteúdo; nesse caso, seria preciso encontrar um modo de ponderar esse resultado atribuindo-lhe um peso menor.

Em termos verbais, algumas palavras e expressões comunicam **intensidade** variada de conteúdo; assim, quando se pretende aferir conteúdos associados por uma pessoa à sua identidade psicossocial, temos frases do tipo: 1) "sempre sou brasileiro"; 2) "às vezes me sinto brasileiro"; e 3) "já esqueci meu passado no Brasil". Poderíamos coligir todos os estratos anteriores sob a denominação geral de formas de expressar identidade nacional, contudo, assim perderíamos a dimensão intensidade que está explicitamente afirmada nas mesmas.

Outra forma de quantificar uma dimensão de conteúdo importante é a **direção.** Já nos primeiros estudos sobre atitudes, esta dimensão recebeu tratamento instrumental em termos de distinguir a orientação em relação a objetos que tendemos a agir, pensar ou sentir de modo desfavorável, neutro ou favorável. Vários instrumentos de Psicologia em cujo construto aparecem conteúdos subjetivos, como os de personalidade, autoconceito, diferencial semântico, entre outros, trabalham com direções bipolarizadas, utilizando escalas de vários graus de intensidade. Assim, na construção de instrumentos para medir intensidade, é necessário saber quais são as formas de expressar os conteúdos opostos, mais facilmente compreensíveis nas respectivas populações e culturas.

A **ordem** de aparição de conteúdo pode constituir um elemento informativo importante para a A.C. Por exemplo, no estudo do fluxo verbal através de associação livre, parece que os primeiros termos associados referem-se às dimensões partilhadas socialmente, tais como estereótipos, que facilitam a comunicação, passando, paulatinamente, para aqueles conteúdos mais pessoais. No estudo de D'Unrug (1974) a respeito da seqüência na produção do discurso, ela chama a atenção para os ritmos, progressões e rupturas de conteúdos para uma compreensão mais completa de uma entrevista.

Ainda outra maneira de organizar quantitativamente a A.C. é a dimensão **espaço-tempo,** que pode ser entendida como o modo de distribuição e de localização dos conteúdos entre si dentro do texto. Ou seja, as dimensões espaço-tempo referem-se aos aspectos gráficos e de *lay-out* (tamanho, superfície do texto, tipografia), bem como ao tempo que se gasta para falar sobre um assunto, por exemplo. Outro modo de tratar a dimensão espaço já mencionada é a co-ocorrência dentro de contexto delimitado de conteúdos, podendo ser associados ou dissociados, conforme o caso. Este tipo de análise complementa as demais, uma vez que a freqüência pode não revelar que em determinados momentos ou lugares não se fala sobre um assunto ou de uma certa maneira.

Ao escolher uma estratégia de codificação e quantificação, o analista delimita suas possibilidades de fazer inferências, as quais poderão se basear, primordialmente, em seu aspecto lógico e/ou estatístico. Sabemos que a maioria das estratégias de codificação e quantificação anteriormente mencionadas referem-se a variáveis dicotômicas ou, no máximo, hierárquicas, impedindo tratamento quantitativo mais sofisticado, como o usado nos testes multivariados. Tal condição, ao invés de ser considerada uma desvantagem para a A.C., deve ser entendida, por um lado, como um desafio e, por outro, como uma necessidade para o tratamento adequado do dado simbólico.

Em termos lógicos, a categorização é uma atividade de classificação de conjuntos de elementos diferenciados segundo características simbólicas intrínsecas ou extrínsecas, podendo ser definidas através de analogias genéricas, em aspectos mais concretos e mais abstratos. Dependendo da condição de produção que o analista pretende focalizar em sua A.C., ele usará uma dimensão ou conjunto delas. O valor da categorização reside precisamente na quantidade de informação que ela é capaz de dar conta a respeito de um fenômeno, mecanismo ou processo existente. Não se trata apenas de generalizar, mas também de oferecer uma nova leitura da realidade com o máximo de objetividade possível. Para tanto, não existem receitas prontas, mas a necessidade de adotar uma postura paradoxal, que inclua a modéstia da observação meticulosa do detalhe e a imaginação generalizadora dos conjuntos sintéticos.

Enquanto conhecimento, a categorização é atividade rotineira de qualquer ser humano; contudo, a categorização em A.C. sistemática se caracteriza pela observação de algumas regras adotadas por vários especialistas:

1) A **exclusão mútua** — exigência de que cada elemento do conjunto analisado seja classificado a partir de categorias específicas, impedindo a ambigüidade e/ou superposição entre as dimensões da realidade. Suponhamos que quiséssemos saber as atitudes de brasileiros a respeito de outros países. Através de um questionário com perguntas abertas, poderíamos classificar as respostas em termos de dimensão: uma dimensão relacionada a esferas da vida ou temas; outra seria a postura global do sujeito em relação a estas esferas, podendo ser favorável, desfavorável ou indiferente. Assim, teríamos duas dimensões sendo analisadas a partir do mesmo material, constituindo sistemas categoriais separados. De todo jeito, a exclusão depende de outra regra que trata da especificidade de cada sistema categorial.

2) **A homogeneidade** — refere-se à expectativa de que a categorização parta de um sistema único de definição e de dimensão da realidade. Assim, os vários elementos classificados devem ter uma relação entre si, mesmo que seja por oposição ou complementação. Tal sistema deve,

nesse sentido, recobrir a totalidade de casos particulares existentes no conjunto de material a ser observado, eliminando o mínimo possível os dados considerados impertinentes ou inclassificáveis, comumente designados pelo termo "outros". De modo que um sistema homogêneo elegante deveria ser capaz de, globalmente, incluir, sem prejuízo, todas as particularidades possíveis em um conjunto de dados, reunidos em função de uma dimensão da realidade de cada vez.

3) **A pertinência** — diz respeito à exigência de que a análise empreendida, em termos de material descrito e categorias definidas, corresponda à problemática de pesquisa e/ou de hipóteses levantadas. Na atividade de pesquisa ou no exercício profissional, trata-se, em geral, de encontrar aqueles índices simbólicos que estão associados ou relacionados à realidade que se pretende conhecer ou modificar. A pertinência da categorização é avaliada a partir da sua capacidade de ¡solar aqueles elementos ou conjunto deles em face de outros existentes que melhor explicam ou modificam os fenômenos e processos em foco.

4) A **objetividade** — regra segundo a qual a categorização deve prover aquelas informações para a classificação do material suficientes para que outro analista independente possa obter os mesmos resultados em pelo menos 70% do material analisado. Contudo, para se garantir a objetividade, o princípio básico, em termos de procedimento, é criar categorias operacionalizáveis, ou seja, passíveis de serem observadas empíricamente de modo manifesto. Sabemos que tal critério foi inicialmente proposto porcomportamentalistas e, posteriormente, por cognitivistas, que tenderam, via de regra, a reduzir a vida simbólica à esfera do comportamento ou ao processo mental individual ou interindividual, preocupando-se mais com o tratamento de informações. Além disso, a fragmentação de fenômenos mais complexos obedeceu também a razões pragmáticas, no sentido de medir com rapidez comportamentos para atingir objetivos diversos. Contudo, não se pode negar que a objetividade é um dos critérios mais importantes, cabendo aos interessados em forjar um trabalho de A.C. mais abrangente da vida humana e social elaborar categorias mais válidas de modo o mais objetivo possível, desde que tal objetividade pretendida não bloqueie a necessidade de dar conta da realidade. Nesse sentido, antes de apresentarmos alguns exemplos de categorização, gostaríamos de enfatizar que o uso de categorização deve obedecer ao critério de minúcia a ser complementado pela criação de tipologias e metassistemas categoriais, que incluam aspectos descritivos mais isolados num nível de abstração mais elevado e/ou que facilitem aformulação de uma teoria mais compreensível que, necessariamente, articule mais de uma dimensão da vida simbólica. A título de exemplo, poderíamos lembrar o trabalho de Piaget sobre o desenvolvimento cognitivo, no qual ele soube descrever, através de índices simbólicos explícitos, as características de cada estágio do mesmo, articulados num metassistema único, que foi, posteriormente,

testado com sucesso em várias populações do mundo. Se ele tivesse simplesmente levantado partes isoladas do fenômeno que estudava, como presença ou ausência de imagens, símbolos e operações em diferentes fases do desenvolvimento humano, sua teoria não existiria.

Na tabela 12-1 apresentamos alguns sistemas categoriais obtidos através de A.C., os quais foram selecionados em função de utilizarem alguns tipos de tratamento quantitativo.

Tabela 12-1 — Temas sobre o morador de Brasília (porcentagens)

Temas	Favoráveis	Neutros	Desfavoráveis
Pessoal	17,96	2,68	7,50
Interpessoal	7,23	-	19,22
Cultura/Educacional	3,75	2,94	13,67
Social/Trabalho	1,87	1,07	6,70
Outros	5,36	-	-
Total	36,17	6,69	47,09

Na Tabela 12-1 temos os resultados de uma A.C. de representações sociais sobre o morador de Brasília, segundo moradores da mesma cidade. Do ponto de vista da análise categorial, foram criados temas para organizar o material obtido através de entrevista aberta, assim como atitudes veiculadas através de significados que os temas apresentavam. Pode-se aplicar um teste de qui-quadrado ou de contingência para verificar a significancia das freqüências para os temas e atitudes. Note-se a significativa concentração do tema **Interpessoal** desfavorável, constituindo-se objeto de pesquisas posteriores.

A Tabela 12-2, a seguir, mostra alguns resultados de um estudo de A.C. para comparar, em termos de freqüência, as dimensões psicossociais mais importantes para a difusão da AIDS entre vários grupos sexuais. Partimos da impressão de que homossexuais masculinos e heterossexuais femininos poderiam ser afetados por uma concepção da interação afetivo-sexual predominantemente heteronômica. Ou seja, estes grupos tenderiam a levar mais em conta as demandas e opiniões do parceiro do que as deles mesmos. Para tanto, elaboramos um questionário que continha perguntas relativas ao uso do preservativo e a alguns aspectos considerados importantes na interação e aplicamos às respectivas populações. Os resultados da aplicação de teste Mann-Whitman nos permitiram confirmar a suposição referida apenas para o grupo homossexual masculino estudado (Souza Filho, 1992).

Tabela 12-2 — Motivos para não usar preservativo

Motivo	Local 1	Local 2	Masculino	Feminino
Acha sem eficácia	0,61	0,36	0,42	0,31
Difícil aquisição	0,68	*0,27	0,28	0,28
Desnecessário	0,88	*1,72	1,78	1,64
Diminui prazer	0,81	*1,52	2,03	*1,10
Por ser incômodo	0,91	1,37	1,75	1,10
Pedido do parceiro	1,10	0,64	0,53	0,75

* Diferenças significativas a $p < 0,05$.

Em estudo correlacionai (Souza Filho, 1988), partiu-se da suposição de que estudantes universitários brasileiros em países desenvolvidos tenderiam a apresentar alguns tipos principais de auto-representação, os quais se organizariam em torno dos conteúdos simbólicos seguintes:

— **identificação com o Brasil** (Ml): "passei ater um sentimento de nacionalidade", "continuo brasileiro";

— **identificação com o** país-anf**itrião** (M2): "sinto-me mais germânico", "continuo admirando alguns dos seus valores (...)";

— **mudança acadêmico-profisslonal** (M3): "(adquiri) rapidez na realização de tarefas", "formação universitária mais sólida", "conhecimento técnico";

— **mudança de ordem geral** (M4): "dividir tarefa com a mulher". "visão ampliada do mundo";

— **mudança sócio-profissional** (M5): "desejo de realizar algum trabalho social", "integrá-lo na realidade e na história da América Latina";

— **mudança individual** (M6): "tornei-me mais frio", "mais esperto", "autoconfiante".

Outros conteúdos relevantes para o mesmo estudo foram:

— **incerteza de reconhecimento no Brasil** (Ia): "ainda não tenho local de trabalho quando voltar ao Brasil", "terei que lutar muito para impor minhas novas idéias ao meu Departamento";

— **incerteza de reconhecimento no país-anfitrião** (lb): "há quatro anos ouço o meu orientador dizer que o que eu faço é ruim, sem interesse, etc", "estou sendo tratado um pouco como pessoa inferior";

— **incerteza de validade de estudos no exterior** (le): "tecnologia inadequada às nossas condições", "vim para cá para servir de mão-de-obra (...)";

— **engajamento em melhorar a situação do Brasil como um todo** (E): "o que é relevante pessoalmente pode não ser relevante para o país como um todo", "brasileiro ajudando o Brasil a procurar um caminho próprio";

— **atitudes em relação ao país-anfitrião (AP);**

— **atitudes em relação ao Brasil** (AB).

Tabela 12-3 — Auto-representações, incertezas e atitudes de estudantes brasileiros em países desenvolvidos (entre parênteses, a significancia)

	Ml	M2	M3	M4	MS	M6	E
Ml					.4905 (<.035)		
M6			.2018 (<.029)				-.1026 (<.088)
Ia			.3090 (<.009)			.2777 (<.014)	
Ib		.8224 (<.000)					.2683 (<.021)
Ie	1.0000 (<.000)						
AP				.1442 (<.034)			-.1464 (<.005)
AB					.4570 (<.015)		.1178 (<.042)

Assim, o cruzamento de elementos de conteúdo permitiu que se constatasse a existência de uma tipologia ou maneiras de enfrentar e de se orientar em face da problemática de ser-brasileiro-estudante-em-país-desenvolvido, apesar de os quocíentes de Pearson terem sido relativamente baixos, podendo ser testada, posteriormente, através de outro instrumento. Mesmo assim, a Tabela-3, acima, indica que a auto-representação de **identificação com o Brasil** tendeu a estar associada à **incerteza de validade de estudos no exterior** e **à mudança sócio-profissional,** enquanto esta, por sua vez, se associou à **atitude favorável em relação ao Brasil. Já mudança acadêmico-proflssional** se associou mais à **mudança individual** e, esta última, à **incerteza de reconhecimento no Brasil.** Além disso, **identificação com o país-anfitrião** tendeu a se associar à **incerteza de reconhecimento no país-anfitrião,** e esta, a **engajamento em melhorar a situação do Brasil.**

Em termos de estratégia, diferentes descrições e tratamentos de dados ora apresentados indicam formas de fazer inferência a partir de

A.C., as quais podem ser interpretadas de acordo com os objetivos e hipóteses do analista.

CONCLUSÕES E PERSPECTIVAS

O princípio geral que orienta o trabalho de A.C. é a necessidade de se dispor de técnicas adequadas para lidar com o fenômeno simbólico humano, de modo sistemático e objetivo. Diante de tal complexidade, que fascina e causa perplexidade há longa data, a tendência científica e profissional geral foi, historicamente, operar por reduções, trazendo um progresso significativo para a Psicologia e disciplinas conexas. Contudo, nota-se uma paulatina ampliação, ao longo do tempo, do número de variáveis e parâmetros considerados importantes para o estudo do simbólico, bem como um crescente questionamento do modo de observar os mesmos, podendo-se afirmar que o campo passou de uma fase "psicológica" para uma "psicossocial" e, hoje, continua procurando integrar contribuições de campos limítrofes, tais como Lingüística e Sociologia, entre outros. Apesar dessa crescente complexificação conceptual e tecnológica existente em A.C. nos dias atuais, permanece vivo entre os que a praticam o interesse pelo dado simbólico natural, aliado aos novos desafios profissionais que emergem continuamente nos diversos contextos sócio-culturais humanos, obrigando aperfeiçoamentos do método e respectivas concepções de trabalho. Nesse sentido, a grande contribuição da A.C. está em oferecer, através de uma reflexão epistemológica a respeito do fenômeno simbólico e de procedimentos e técnicas de análise, as condições que o determinam em termos de produção, transmissão e recepção.

Na cadeia de etapas por que passa a produção do conhecimento científico e profissional em ciências humanas, a A.C. situa-se em sua fase primordial, básica. Pois, mesmo partindo de algum quadro hipotético e categorial preestabelecido, só se pode considerar A.C. o contato "direto" do analista com a realidade simbólica natural a ser investigada, como é o caso de material obtido através de entrevistas, conversas, filmes, desenhos, escritos, entre outros. Nesse sentido, na formação acadêmica atual, tem-se negligenciado o papel desse tipo de instrumental para a construção de questionários e intervenção. Nos países em desenvolvimento, tal necessidade parece ser maior, uma vez que boa parte da pesquisa decorre de teorias e métodos produzidos e testados, primordialmente, em outros contextos a A.C., devendo esta ser introduzida, portanto, como etapa de validação dos mesmos. Contudo, acreditamos que a A.C. facilite a emergência de novas abordagens teóricas e práticas em geral.

REFERÊNCIAS BIBLIOGRÁFICAS

ADORNO. T.W. *The authoritarian personality.* New York: Harper, 1950.

BARDIN, L. *L'analyse de contenu.* Paris: P.U.F., 1977/1992.

BARTHES, R. *Mythologies.* Londres: Jonathan Cape, 1972.

--------------- . *Le système de la mode.* Paris: Seuil, 1967.

BERELSON. B. *Content analysis in communications research.* New York: Hafner, 1952.

---------------. Content analysis. In: LINDZEY,G.(Ed.). *Handbook of social psychology.* Cambridge (MA): Addison-Wesley, 1954.

CARTWRIGHT, D.P. Análise do material qualitativo. In: FESTINGER, L., KATZ, D. (Orgs.). *A pesquisa na psicologia social.* Rio de Janeiro: Fundação Getúlio Vargas, 1974.

CHOMBART DE LAUWE. M.J. *Enfants de l'image.* Paris: Payot. 1979.

DE ROSA, A.S. The social representations of mental illness in children and adults. In: DOISE, W., MOSCOVICI, S. (Eds.). *Current issues in European social psychology.* Cambridge: Cambridge University Press, 1987. v.2.

DOISE, W., CLEMENCE. A., LORENZI-CIOLDI, F. *Représentations sociales et analyses de données.* Grenoble: Presses Universitaires de Grenoble, 1992.

D'UNRUG, M.-C. *Analyse de contenu et acte de parole.* Paris: Delarge: Ed. Universitaires, 1974.

DURKHEIM, E. *Suicide,* a study in sociology. Glencoe: Free Press, 1951.

FISHBEIN, M., AJZEN.. I. *Belief, attitude, intention and behavior,* an introduction to theory and research. Cambridge (MA): Addison-Wesley. 1975.

GHIGLIONE, R.. BLANCHET. A. *Analyse de contenu et contenus d'analyses.* Paris: Dunod, 1991.

HOLSTI. O.R. Content analysis. In: LINDZEY. G., ARONSON. F. (Eds.). *Handbook of social psychology.* Cambridge (MA): Addison-Wesley. 1969.

KATZ, E. The two steps of communication. In: SCHRAMM. W. (Ed.). *Mass communication.* Glencoe: Free Press, 1960.

KRIPPENDORFF, K. *Content analysis.* Londres: Sage, 1980.

LASSWELL, H.D., LEITES, N. *The language of politics:* studies in quantitative semantics. New York: G.Stewart, 1949.

LAZARSFELD, P.F.. BERELSON, B„ GAUDET, H. *The people's choice.* New York: Columbia University Press, 1948.

MAINGUENEAU. D. Nouvelles tendances en analyse du discours. Paris: Hachette, 1987.

McCLELLAND. D.C.. FRIEDMAN. CA. Across-cultural study of the relationship between child training practices and achievement motivation appearing in folk tales. In: SWANSON. G.E., NENCOMB. T.M.. HARTLEY, E.H. (Eds.). *Readings in social psychology.* New York: Holt 1953.

MOSCOVICI, S. Communication process and the properties of language. In: BERKOWITZ, L. (Ed.). *Advances in experimental social psychology.* New York: Academic Press, 1967. v.3.

----------------. *La psychanalyse:* son image et son public. Paris: P.U.F., 1961/1976.

MUCCHIELLI, R. *L'analyse de contenu de documents et des communications.* Paris: Los Libraries Techniques, 1979.

OSGOOD, CE. The representational model. In: POOL, S. (Ed.). *Trends in content analyses.* Urbana (IL): University of Illinois Press, 1959.

OSGOOD, CE., SUCI,G.J.,TANNENBAUM,P.H.77îe measurement ormean/ng. Urbana (IL): University of Illinois Press, 1957.

OSGOOD, CE., WALKER, G. Motivation and language behavior: a content analysis of suicide notes. *Journal of Abnormal and Social Psychology,* v.59, p.58-67, 1959.

PÉCHEUX, M. *Analyse automatique du discours.* Paris: Dunod. 1966.

POTTER, J., WETHERELL, M. *Discourse and social psychology.* Londres: Sage. 1987.

SOUZA FILHO, E.A. de. Alguns pressupostos teóricos e metodológicos do estudo de representações sociais. SIMPÓSIO BRASILEIRO DE PESQUISA E INTERCÂMBIO CIENTÍFICO EM PSICOLOGIA, 3, 1990. *Anais...,* p.384-387. [S.l.: s.n.), 1990.

---------------. *Heterossexuais e homossexuais no Rio de Janeiro no contexto da AIDS.* No prelo.

---------------. *Personagens da revista Sabrina:* uma análise de conteúdo. No prelo.

---------------. Representações sociais de Brasília. *Psicologia e Sociedade.v.9,* p. 153-159,1990/ 1991.

---------------. Universitários brasileiros no exterior: uma análise psicossocial. *Ciência e Cultura,* São Paulo, V.40, p.559-565, 1988.

WEBER, R.P. *Basic content analysis.* New York: Sage, 1985.

WICKLUND, R.A.. DUVAL, S. *A theory of objective self awareness.* New York: Academic Press, 1972.

CAPÍTULO 13

MEDIDA PROJETIVA

Jurema Alcides Cunha, Maria
Lúcia Tiellet Nunes
Pontifícia Universidade Católica do Rio Grande do Sul — PUCRS

O título proposto para este capítulo envolve aparentemente uma inconsistência lógica, porque, embora o manejo de testes projetivos possa ser intentado em termos quantitativos, a maioria dos autores que defendem o seu uso o faz visando à exploração de aspectos dinâmicos da personalidade que adquirem significado sob a ótica de um referencial teórico ao qual há difícil acesso via psicométrica.

A questão subentendida pode ser caracterizada como polêmica. É verdade que o termo medida pode ser tomado sob várias acepções e, se o considerarmos mais literalmente como mensuração, o tema pode ser abordado, demonstrando que, mesmo nas chamadas técnicas projetivas, o material identificado como projetivo só pode ser interpretado com cautela e dentro de "limites realísticos", autorizados por alguma forma de quantificação que proporcione um grau razoável de certeza (Exner, 1989, p.534). Já num sentido figurado, o vocábulo medida pode ser entendido como "meio de comparação e julgamento" (Ferreira, 1986 p.l 109) e, então, poder-se-ia enfatizar o método de enfoque projetivo "diferente daquele que fundamenta os testes psicométricos" (Anzieu, 1981, p.16). Mas aqui, novamente, se pisa num terreno movediço. Por um lado, **projetivo** é um adjetivo que é utilizado, no campo do psicodiagnóstico, para qualificar técnicas cujos estímulos e instruções potencialmente pretendem eliciar no sujeito o processo de projeção, ainda que possam também facilitar respostas que representam simplesmente o produto do manejo lógico do material estímulo. Assim, há autores que preferem denominar as técnicas projetivas de indiretas (Vane, Guarnaccia, 1989). Por outro lado, o termo se aplica a determinadas respostas que, de forma alguma, são exclusivas a testes projetivos, podendo emergir no contexto estruturado da testagem com instrumentos psicométricos. Como escreve Shneidman (1965, p.498), em seu texto clássico sobre técnicas projetivas, este é um assunto complexo cujo "conceito e a própria denominação colocam algumas questões fundamentais que demandam reflexão" — em grande parte, a seu ver, a

partir de uma confusão entre o conceito de projeção e técnicas projetivas, que envolvem palavras "que soam similarmente, mas que têm significações diferentes" (Shneidman, 1965, p.501).

Portanto, neste capítulo, procuraremos examinar alguns pontos discutíveis implícitos nas questões em foco e subentendidos por uma expressão pouco específica — medida projetiva —, que parece ter potencial para encontrar sentido próprio no contexto da avaliação psicológica, como uma outra dimensão para o entendimento da personalidade, se considerarmos o uso clínico extensivo das técnicas projetivas hoje em dia.

CONCEITO DE PROJEÇÃO E TÉCNICAS PROJETIVAS

Ainda que não imunes a outras correntes de pensamento, a maioria dos autores, associados a técnicas projetivas durante este século, esteve, de uma forma ou de outra, empenhada em reflexões sobre pressupostos psícanalíticos e, particularmente, preocupada com as possiveis relações entre uma psicologia projetiva e o conceito de projeção freudiano.

Podemos situar as raízes desses movimentos em 1894, quando Freud começou timidamente a esboçar o conceito de projeção ao tentar explicar uma estratégia do psiquismo para lidar com a angústia frente a uma excitação interna não controlável, ao comportar-se "como se estivesse projetando tal excitação para fora" (Freud, 1987b, p. 109), apesar de que, provavelmente, só tenha chegado a usar o termo projeção em obra publicada em 1896, conforme comentário de Stachey (Freud, 1987c). Nos anos subseqüentes a 1894, Freud continuou a examinar os processos utilizados para evitar a incursão na consciência de representações inconscientes, em dois artigos sobre mecanismos de defesa (Freud, 1987c, 1987d) e no Manuscrito H, enviado a Fliess com carta de 24 de janeiro de 1895 (Masson, 1986, p.105-112). Especialmente neste último, discutiu em minúcia o mecanismo típico da paranóia, cujo objetivo seria o de "defender o eu de uma idéia insuportável, que é projetada no exterior" (id.ibid., p. 108-109), definindo suas diferenças em relação a outras defesas. Também, discorreu brevemente sobre o que constituiria uma projeção normal e como seria o uso inadequado desse mecanismo psíquico. Em 1911, chegou a uma explicação mais clara da projeção, como a substituição de percepções internas por externas, através de uma tática de atribuição dos próprios sentimentos e impulsos a pessoas e objetos externos (Freud, 1969). No ano seguinte, em *Totem e Tabu,* ampliou o conceito, ao admitir que projeção é um processo que nem sempre emerge como defesa ante um conflito, considerando-O um mecanismo primitivo, que pode influenciar "nossas percepções sensoriais, e que, normalmente, desempenha um papel muito grande na determinação da forma que toma nosso mundo

exterior", de modo que "as nossas percepções internas de processos emocionais e de pensamento podem ser projetadas para o exterior da mesma maneira que as percepções sensoriais" (Freud, 1974b, p.86). E, mais adiante, fez uma afirmação muitas vezes retomada como pressuposto teórico da psicologia projetiva, ao dizer que, assim, "estamos reconhecendo a existência de dois estados — um em que algo é diretamente fornecido aos sentidos (ou seja, está **presente** neles) e, ao lado deste, outro, em que a mesma coisa é **latente,** mas acaba de reaparecer—ou, essencialmente, estamos reconhecendo a coexistência da percepção e da memória ou, em termos mais gerais, a existência de processos internos **inconscientes,** ao lado dos **conscientes"** (id. ibib., p. 117).

Como chamam a atenção Laplanche e Pontalis (1968), Freud retomou o tema numa série de outros pontos de sua obra. Não obstante, segundo Exner (1989), é em *Totem e Tabu* (1974b) que ele explicou a projeção de forma mais aplicável ao conceito implícito em testes e respostas, que se qualificam por seu sentido projetivo, adotado por Murray, para esclarecer o fenômeno observado na interpretação que o sujeito faz no TAT, ao integrar "experiências passadas e necessidades presentes", expressando conteúdos de nível consciente e inconsciente (Murray, 1951, p.38).

Murray, conforme Bernstein (1951), admitiu que o conceito de projeção emprestava peculiaridade à sua técnica, estabelecendo uma diferenciação tríplice no processo, como projeção externalizadora, defensiva e expressiva. No entanto, o sentido que deu à projeção, referido acima, fez com que Exner (1989) recapitulasse o ponto de vista de Shneidman (1965), ao concordar com Lindzey (1961) em sua afirmação de que Murray foi o responsável, em 1938, pelo estabelecimento do elo de ligação entre o conceito de projeção e as técnicas projetivas. Esse vínculo tem sido atribuído, inclusive por Bernstein (1951), à posição defendida por Franck, em sua obra publicada em 1939. Não obstante, como Lindzey (1961) salientou, a popularização das expressões "técnica projetiva" e "método projetivo", sem dúvida, ocorreu pela influência do trabalho de Franck, ainda que, paradoxalmente, este autor tenha demonstrado completo desinteresse pelo referencial freudiano.

Ele simplesmente aproveitou "o termo, numa rudimentar analogia espacial", para designar "a tendência geral do indivíduo para revelar porções significativas de sua personalidade, quando está estruturando, de forma livre, objetos do mundo exterior" (Lindzey, 1961, p.37). Ao explicar tal conceito, segundo Shneidman (1965, p.501), enfatizou o caráter holístico da personalidade e a "natureza idiossincrásica da percepção humana". Mas, como suas idéias assumiram a liderança na psicologia projetiva, sua posição constituiu o ponto de partida para as crescentes dificuldades, nesse campo, de harmonizar os pressupostos freudianos com o conceito implícito nas técnicas projetivas. Todavia, observa-se que a maioria dos

adeptos do enfoque projetivo tem se notabilizado por seus esforços para definir um embasamento teórico de cunho psicanalitico, sofrendo, entretanto, outras influências de correntes de pensamento.

Abt, por exemplo, procurou conciliar o referencial teórico da psicologia da **gestalt** com pressupostos psicanalíticos, ao analisar a natureza e a função da percepção. Da ação da percepção, dependeriam todos os métodos projetivos, sustentando o autor que mecanismos de defesa, dos quais o mais importante seria a projeção, atuariam sobre a percepção, permitindo que ela "desempenhe um papel significativo no processo de homeostase psicológica" (Abt, 1967, p.46), ao possibilitar o manejo adequado da ansiedade. Como comenta Exner (1989), Abt é o exemplo de um autor que, ao lado de outros neste século, centrou sua fundamentação teórica no caráter defensivo da projeção freudiana, ainda que lhe emprestando nova roupagem e enfatizando seu sentido homeostático.

Bellak (1967) inicialmente adotou um ponto de vista praticamente idêntico ao de Abt. Mais tarde, com base em resultados de investigações, rejeitou a conotação indiscriminada de defesa no conceito de projeção. Retomou a obra de Freud, mais especificamente sua discussão referente à projeção, em Tofem e *Tabu* (Freud, 1974b), detendo-se em sua pressuposição da coexistência de processos conscientes e inconscientes, para explicar a influência de percepções passadas sobre a percepção atual dos estímulos. Julgando que a designação de percepção, na época, podia estar comprometida com uma concepção não holística da personalidade, optou pelo termo **apercepção** (antes já utilizado por Murray), para se referir a "uma interpretação (dinamicamente) significativa que o organismo faz de uma percepção" (Bellak, 1967, p.27). Caracterizou, então, a apercepção pela carga subjetiva de conteúdos da experiência passada do indivíduo, que impregnam sua percepção atual, provocando um fenômeno chamado de **deformação aperceptiva.** Assim, a partir de uma hipótese de trabalho, subentendida pelo conceito de apercepção, diferenciou "graus de deformação aperceptiva" (Bellak, 1979, p.21), que seriam:

a) projeção invertida, em que a deformação aperceptiva encontrada em respostas ao teste é mais severa, por um processo de defesa, tal como Freud descreveu no caso Schreber (Freud, 1969b);

b) projeção simples, em que a deformação aperceptiva se baseia em conteúdos mnêmicos anteriores, que podem se transferir à nova situação, não tendo necessariamente significado clínico;

c) sensibilização, em que a deformação aperceptiva se justifica, porque a situação atual mobiliza um modelo de resposta aprendido anteriormente em outra situação com conotações semelhantes, mas que é inadequado no momento;

d) percepção autistica, em que a deformação aperceptiva acontece a partir de lembranças perceptuais associadas com impulsos básicos

primitivos, eventualmente considerada como um processo de projeção simples; e

e) externalizaçáo que, ao contrário dos anteriores, "dos quais o indivíduo comumente não está consciente" (Bellak, 1967, p.29), é um processo que pode ser considerado pré-consciente, em vista de seu "fácil acesso à consciência" (Bellak, 1979, p.25).

Conforme Bellak (id. ibid.), enquanto Murray distinguía subtipos de projeção, analisando seu conteúdo específico, ele se preocupou em definir "o grau de severidade ou complexidade" ou, ainda, "o grau da inconsciência relativa da deformação". Porém, não se deteve apenas no comportamento projetivo — que é o que envolve a deformação aperceptiva —, mas examinou o papel do comportamento adaptativo — que tem um caráter perceptivo, porque depende fundamentalmente dos elementos reais e objetivos do estímulo — e, também, do comportamento expressivo — que reflete as diferenças individuais que determinam não o conteúdo da resposta, mas sim suas características (Bellak, 1979), o seu estilo (Wolff, Precker, 1976), afirmando que esses comportamentos "sempre coexistem" (Bellak, 1979, p.26). Além disso, também se diferenciou da posição de Murray, quando, sob a influência das idéias de Hartmann (1968), sugeriu que o ego constitui uma variável interveniente entre o conteúdo latente e o manifesto, chegando a abordar a questão da regressão a serviço do ego na produção imaginativa, tema retomado mais tarde em outro contexto (Bellak, Hurvich, Gediman, 1973) e melhor explorado, em relação às técnicas projetivas, por Schafer (1954, 1978) e Holt (1970).

Cattell (1976, p.89) considerou errônea a ênfase dada pelos pioneiros da psicologia projetiva ao conceito psicanalitico de projeção, salientando que outros mecanismos de defesa poderiam contribuir para que aspectos dinâmicos "rompessem as barreiras da consciência", pelo que propôs que os instrumentos para sua investigação se chamassem **testes de dinâmica de defesas do ego** ou, simplesmente, **testes de dinâmica** ou **testes de defesa.** Esses testes deveriam apresentar a qualidade metodológica dos chamados "testes objetivos", uma vez que avaliam diferenças individuais numa atividade cognitiva.

Uma série de fatores poderia ser responsável pela percepção diferenciada individualmente, enquanto comparada com uma "percepção normal". O estudo de tais diferenças poderia ser feito através de testes de apercepção ou de percepção errônea, que permitem a investigação "de distorções ocasionadas por necessidades dinâmicas" e, portanto, não se referem apenas "às suscitadas por efeitos cognitivos, associados à inteligência ou à memória", por exemplo (id. ibid., p.131-132). Essas distorções podem resultar de "estados emocionais transitórios", de desejos ou, ainda, de "traços dinâmicos permanentes" (id. ibid., p.96), que hoje seriam caracterizados como egossintônicos ou não. Quanto à percepção errônea,

esta é explicada a partir de um conceito amplo de projeção. Cattell (1976, p.99) discute as diferentes modalidades de atuação e operacionalização da percepção errônea e aplica essa classificação tanto para uma "inferência ingênua", para uma distorção, em que o sujeito transforma as percepções "em algo semelhante ao objeto de seus desejos", como para o reflexo de uma necessidade de tornar a realidade externa compatível e consistente com aspectos subjetivos interiores ou, ainda, para um mecanismo defensivo, caracterizado pela atribuição de aspectos interiores inaceitáveis a objetos do mundo exterior. Possivelmente seja em razão desse sentido amplo que Exner (1989, p.521) afirma que Cattell considerou a projeção simplesmente "como um processo que fornece informações sobre a personalidade".

Rapaport (1965) aceitou a hipótese projetiva como embasamento dos testes projetivos, mas deixou claro que o conceito de projeção adotado não pressupunha a atribuição de sentimentos e impulsos a pessoas do mundo externo nem tinha o caráter de mecanismos de defesa. Admitiu que a projeção pode ocorrer em técnicas diretas e indiretas de avaliação psicológica, porque as manifestações do comportamento do indivíduo são reveladoras de sua personalidade. Porém, em sua opinião, os instrumentos que podem ser considerados mais especificamente como projetivos são aqueles "em que o sujeito, ativa e espontaneamente, estrutura material não estruturado e, ao fazê-lo, revela seus princípios estruturadores, que são os princípios de sua estrutura psicológica" (Rapaport, Gill, Schafer, 1972, p.225).

Rapaport (1978) aceita o que chama de modelo psicológico da psicanálise porque o entendimento das respostas dadas a uma técnica projetiva — que refletem a estrutura psicológica do sujeito — "implica uma teoria de personalidade que pressuponha que grande parte da estrutura psicológica não é conscientemente experienciada por ele", de forma que estão em jogo motivações inconscientes e, assim, é necessária "uma teoria de personalidade que pressuponha a existência dessas motivações e as explique" (Rapaport, Gill, Schafer, 1972, p.228). Não obstante, utiliza conceitos freudianos de forma crítica, porque "a personalidade se manifesta (nos testes projetivos) através de um processo de pensamento ou através do produto desse processo", e, portanto, "a exploração consistente dos testes projetivos é uma exploração dos processos de pensamento" e sua psicologia é "parte da psicologia do ego" (id. ibid., p.229). A psicologia projetiva, pois, lida fundamentalmente com processos que ocorrem no ego, supondo-se, entretanto, a possibilidade da emergência de material oriundo do inconsciente nesses processos, em especial quando o pensamento carece de organização. Em conseqüência, é necessário o referencial freudiano para o entendimento desse material, mas são básicos os conceitos da psicologia do ego para a compreensão dos processos de pensamento e dos mecanismos de defesa do ego.

Neste ponto, parece oportuno repetir a pergunta levantada por Murstein e Pryer, em 1959: "Quantos tipos de projeção existem?" Em sua ótima revisão, para a época, da literatura existente, concluíram que há cerca de "quatro categorizações possíveis do conceito de projeção: 'clássica', 'atributiva', 'autistica' e 'racionalizada' (Murstein, Pryer, 1959, p.353). A primeira corresponderia ao conceito inicial de Freud (1987b), admitindo, porém, que também traços favoráveis pudessem ser projetados no mundo exterior. A projeção atributiva seria a mais amplamente utilizada no campo da personalidade, como tática de atribuição, nos termos descritos por Freud (1974b) em *Totem e Tabu.* Já a projeção autistica ocorre quando a percepção sofre forte influência das necessidades do indivíduo, no sentido de que "os aspectos formais do objeto percebido são modificados de modo a se tornarem consistentes com a necessidade" (id. ibid., p.354). A projeção racionalizada envolveria um processo que se mantém inconsciente, como no caso da projeção clássica, mas o comportamento é consciente no indivíduo, que o justifica através de uma racionalização. Conforme esses autores, este tipo de projeção foi considerado uma projeção complementar por Allport, uma projeção do recíproco por Piaget, sendo que Van Lennep a definiu não exatamente como projeção, mas como seu correlato.

Após a discussão dessas categorizações possíveis de projeção, conforme a utilização do conceito por autores variados, Murstein e Pryer passaram a examinar pesquisas desenvolvidas sobre o tema, no período entre 1936 a 1957, analisando um total de 45 trabalhos. Deixando de lado alguns, em que foi difícil chegar a conclusões em face dos "métodos variados de mensurar o conceito" (id. ibid., p.358), suas críticas, em geral, se centraram na definição operacional de projeção, afirmando que é necessário que esta "não se afaste da significação psicológica aceita do conceito" (id. ibid., p.368). Por outro lado, a partir da revisão da literatura efetuada, propuseram que a definição de projeção se limitasse "à manifestação de comportamento do indivíduo, que indique algum valor ou necessidade emocional do mesmo", portanto, excluindo "componentes fisiológicos ou cognitivos" e acrescentando que "tal comportamento pode variar no grau de defensividade, dependendo do contexto situacional e da personalidade do que percebe" (id. ibid., p.370).

Exner (1989) também desenvolveu pesquisas sobre a projeção, afirmando que esta não é uma questão simples, em vista da própria complexidade das operações psicológicas que ocorrem durante o processo de resposta a estímulos, como os do Rorschach. Por outro lado, salienta que não é possível partir da pressuposição de que o sujeito, frente a um campo de estímulos não estruturado, impregna de elementos subjetivos a sua percepção atual, pois, se isto fosse verdadeiro como todas as manchas de tinta do Rorschach envolvem características de escassa estrutura e ambigüidade, todas as suas respostas deveriam conter material projetivo.

O mesmo autor faz uma análise das operações percepto-cognitivas que se sucedem no processo de resposta, demonstrando que, embora o processamento de informações seja muito rápido (1983), é possível identificar três fases, ainda que não totalmente distintas entre si (1989), como são apresentadas na Tabela 13-1.

Tabela 13-1 — Fases e operações do processo de resposta

Fase I	Fase II	Fase III
1. Codificação do campo de estímulos	3. Reexploração do campo para refinar respostas potenciais	5. Seleção final das respostas potenciais restantes
2. Classificação da imagem codificada e de suas partes em respostas não usá-veis ou indesejáveis, por ordenação ou censura, na comparação entre si	4. Rejeição de respostas potenciais	6. Articulação da resposta selecionada

Fonte: Exnei, 1080, p.522.

A complexidade das operações psicológicas envolvidas no processo de resposta leva a certas considerações. Por exemplo, não se pode ingenuamente afirmar que, como se tratam de meras manchas de tinta, o fato de o sujeito dar a resposta de morcego à Lâmina I implicaria um afrouxamento do juízo crítico, numa aplicação estrita do conceito de regressão a serviço do ego, que será discutido mais adiante. Além das próprias instruções serem suficientemente flexíveis, o número de sujeitos que dá essa resposta, com base no contorno da mancha, torna claro que "o exemplo mais óbvio da realidade da mancha e de seu impacto sobre o processo de resposta é o escore do determinante F (Peterson, Schilling, 1983, p.268). Reconhece-se, portanto, que elementos do estímulo tendem a compelir certas respostas, "limitando ou inibindo outras classificações" (Exner, 1989, p.522), como se evidencia claramente nas respostas populares. Assim, resultados de pesquisa de Exner e Martin, citados pelo primeiro, demonstraram que, com a exclusão das projeções Dd 34 da mancha da Lâmina I, nenhum sujeito forneceu a tradicional resposta de morcego e apenas 6% produziram a resposta de animal alado. Em face disto, é possível concluir que "as características críticas do estímulo criam limitações na amplitude de respostas que podem ser geradas, sem restringir as propriedades ou realidades do campo" (id. ibid., p.524). Por outro lado, outras características da mancha, além da forma, podem influenciar na escolha das respostas potenciais alternativas. Então, na Lâmina I, o sujeito "mais convencional, preciso e conservador em suas operações decisorias" valorizará o determinante cor e responderá que se trata de um morcego, rejeitando a resposta borboleta, que implicaria a desconsideração "da incongruência do colorido" (id. ibid., p.522). Deste

modo, embora não se possa negar a possibilidade de o sujeito incluir material projetado em sua tradução do estímulo, estudos deteste e reteste realizados pela equipe de Exner revelaram que a seleção final das respostas verbalizadas dependia basicamente de dois fatores — a organização psicológica do sujeito e o seu estado psicológico.

Já as projeções que, segundo Exner, se verificam conforme os processos descritos por Freud e Murray, podem ocorrer em qualquer das três fases do processo de resposta.

Na fase I, os elementos de realidade assumem enorme importância, diminuindo as possibilidades de projeção. Entretanto, eles podem "ser violados ou ignorados" na percepção. Se o fato não resulta de distorções de ordem neurofisiológica, é possível pressupor que decorra "de alguma forma de mediação cognitiva", em que aspectos psicológicos internos modificam "uma tradução do campo orientada para a realidade" (Exner, 1989, p.528).

Nas fases II e III, em que "cada resposta potencial é revisada, julgada e refinada", há condições favoráveis, "em que o sujeito pode embelezar a resposta" (id. ibid., p.529). Mesmo assim, se admite que ela possa representar uma superelaboração de atributos do estímulo. Para testar essa pressuposição, foram realizados estudos com sujeitos que apresentavam escores altos em escalas que mediam certos traços de personalidade, como cooperação, agressão ou morbidez. Foi verificado que eles tendiam a dar, no Rorschach, respostas que envolviam conteúdo idêntico, em número apreciável, embora não de modo exclusivo. Noutro estudo, foram propostas respostas alternativas para determinadas localizações nas dez manchas. Os sujeitos, diferenciados por certos traços de personalidade, tendiam a selecionar, entre tais respostas alternativas, a mais compatível com suas disposições internas. Outra pesquisa demonstrou que, mesmo que as respostas do sujeito possam refletir influências sobre a expectativa de determinados conteúdos, ainda é possível registrar a importância dos efeitos de "disposições internas, criadas por características potentes da personalidade (por exemplo, necessidades, interesses, aptidões, conflitos, etc.)" (id. ibid., p.534).

Em resumo, fica evidente que, embora cada resposta constitua "uma representação indireta das operações psicológicas do sujeito", de forma alguma isto "significa que cada resposta ou partes de cada resposta sejam produtos de projeção" (id. ibid., p.520). Esta é apenas possível, não obrigatória.

Tais considerações, feitas a partir de estudos com o teste de Rorschach, são aplicáveis, em diferentes graus, à questão das respostas a outras técnicas. Por exemplo, o CAT é tido como uma técnica projetiva utilizável a partir dos três anos de idade (Bellak e Bellak, 1981; Hirsch,

Verthelyi, Menéndez de Rodríguez, 1979; Anzieu, 1981). No entanto, investigações feitas com crianças de idade pré-escolar, utilizando um sistema de escore (Cunha, Nunes, Werlang, 1990; Cunha, Werlang, 1991), demonstraram que, até os 5 anos, predominam respostas de caráter não aperceptivo — descrições ou enumerações —, por não conseguir a grande maioria desses pré-escolares ultrapassar a simples percepção dos elementos de realidade do estímulo, portanto, sem evidências de projeção. Já com o aumento gradual da idade, verifica-se que surgem, aos poucos, tentativas de interpretação que, segundo Hoar e Faust (1962, p.245), existe, quando "sentimentos dos sujeitos são descritos, estados psicológicos são inferidos, atividades não representadas nas figuras têm lugar". Ora, é lógico pensar que, somente quando as respostas são de cunho interpretativo, existe alguma probabilidade da ocorrência de projeções, sempre ressalvada a possibilidade de constituírem sobre-elaborações de atributos do estímulo, principalmente quando são elegidos temas freqüentemente eliciados.

Antes, parece ser possível pressupor que, não só no Rorschach, mas em outras técnicas ditas projetivas, se verifique uma série de operações psicológicas no processo de resposta, com a ocorrência, numa primeira fase, de um escrutínio do campo de estímulos e uma classificação dos elementos de realidade do mesmo. Estudos realizados e a revisão da literatura possibilitam que se afirme que o processo — que se desenvolve e até a elaboração da resposta final (seja esta uma história, um desenho ou, mesmo, uma resposta à apresentação de uma mancha de tinta) — parece depender fundamentalmente da "maturidade para cumprir a tarefa" (Cunha et al., 1993, p.264) e de condições de integridade das funções do ego (Montagna, 1989), que podem levar ao embelezamento (Exner, 1989) ou, eventualmente, à deterioração do produto final, julgados esses em termos de parâmetros da realidade.

Portanto, de forma subjacente ao desempenho observável na testagem e de acordo com as expectativas implícitas nas instruções propostas pelas diferentes técnicas, pode-se afirmar, com Rapaport (1978, p.134), que "está o processo do pensamento" naturalmente consubstanciado por outros processos cognitivos. O conjunto de operações psicológicas que se estabelece explica o manejo lógico do material estímulo, que ancora o desempenho aos elementos de realidade, numa tarefa que envolve sempre a solução de problema, que exige "uma adaptação a estímulos exteriores estabelecidos, quer dizer, (que) põe em jogo a **função de realidade"** (Rorschach, 1948, p. 120).

Nessas circunstâncias, a emergência de material projetivo, embora possível, não é obrigatória, nem tão provável ou freqüente como originalmente se supunha. Não obstante, o foco desta discussão envolve uma tentativa de deslindar o que há de projetivo nas técnicas projetivas. Por outro lado, há fatos observáveis que não podem ser descartados e exigem

reflexão. Assim, em protocolos de indivíduos obsessivo-compulsivos, se constata uma aderência rígida aos elementos de racionalidade e o controle estrito na formulação das respostas para se adequarem à realidade. Já em material produzido por indivíduos com personalidade *borderline* ou em muitos exemplos de respostas de esquizofrênicos, podem ser destacados excertos compatíveis com modos arcaicos, primitivos, mágicos de raciocínio, que contrastam extraordinariamente com as regras comuns da lógica.

Para o entendimento dessa variabilidade que se insinua em produtos de testagem, é importante, de um ponto de vista teórico, a consideração do **conceito de regressão a serviço do ego,** "baseado na possibilidade de, com o afrouxamento de controles, haver um movimento regressivo ao longo de um *continuum* entre o processo secundário e o processo primário" (Werlang, Cunha, 1993, p.125).

Freud, principalmente em duas de suas obras (1974a, 1977), mencionou a possibilidade de um tipo de pensamento característico do inconsciente aparecer ao nível da consciência nos chistes, nos sonhos e na psicose. Sería o caso da regressão formal, em que "métodos primitivos de expressão e de representação tomam o lugar dos métodos habituais" (Freud, 1987a, p.501), fenômeno que, posteriormente, foi denominado por Kris como regressão a serviço do ego, conforme referem vários autores. Trata-se de uma mudança no processo do pensamento (Holt, 1970).

Na realidade, Rapaport (1962) sugere que Freud associou os conceitos de processo primário e secundário, respectivamente, aos dois princípios de funcionamento mental. De fato, encontra-se em sua obra que **fantasia** é uma atividade "que foi liberada do teste de realidade e permaneceu subordinada ao princípio de prazer", ao passo que o **pensamento** é explicado como uma função do aparelho mental, coibida de descarga motora, desenvolvida pela "apresentação das idéias", dirigida "para as relações entre impressões de objetos" e que, geneticamente, "não adquiriu outras qualidades perceptíveis à consciência até haver-se ligado a resíduos verbais" (Freud, 1969a, p.281).

Assim, o processo primário, observado em modos de pensamento primitivo ou desorganizado, desconsidera a realidade, regido pelo princípio do prazer, mas tem propriedades formais que permitem o seu entendimento. Já o processo secundário apresenta todas as características do pensamento consciente de uma pessoa normal e civilizada (Holt, 1970).

Tais construtos, criados por Freud, são úteis em seu sentido conceptual, mas representam uma abstração que realmente não se apresenta por si. Antes, pode-se imaginá-los como polaridades, entre as quais pressupõe-se um *continuum,* ao longo do qual o pensamento flui, com "todos os graus de tradição" (Schafer, 1978, p.89), podendo a regressão que ocorre adquirir um caráter adaptativo, se está a serviço do ego, ou não.

No interesse do ego, há uma redução de controle que possibilita que o pensamento se libere das regras rígidas do processo secundário, abastecendo-se de recursos e fenômenos mais comumente utilizados nos sonhos. Há afrouxamento, mas não perda de controles, de modo que é facilmente reversível, culminando "em fases progressivas, em que os produtos do movimento regressivo são elaborados ou sintetizados" de uma forma "orientada para a realidade objetiva" (Schafer, 1978, p.87), resultando numa criação artística ou numa resposta bem organizada e embelezada a técnicas projetivas.

Mesmo a serviço do ego, o movimento regressivo pode se mostrar vulnerável a conflitos anteriores, não completamente elaborados, havendo indícios que facilitam seu entendimento psicodinâmico. Mas, quando as fantasias são muito perturbadoras, costumeiramente são mobilizados sinais afetivos de alerta que diminuem os perigos que ameaçam o êxito da reversão do processo, sendo que esses são mais sérios quando as funções do ego são mais frágeis, menos autônomas para neutralizar impulsos mais primitivos.

Portanto, se a possibilidade de se efetuar o movimento regressivo depende de certa flexibilidade do ego, que permite um afrouxamento das regras mais rígidas do processo secundário, também a reversão do processo, para propósitos adaptativos, é possibilitada por determinadas condições do ego.

Como as regressões normais do desenvolvimento infantil, que são temporárias e reversíveis, são processos que constituem — como afirma Anna Freud (1971, p.95), concordando com Spitz — "respostas úteis à tensão de um determinado momento", da mesma maneira se pressupõe que, numa situação de testagem, o processo ocorre quando os estímulos provocam uma mobilização afetiva muito intensa. Sendo o movimento regressivo a serviço do ego, há fatores facilitadores da reversão do processo, principalmente a existência de um sentido primário de identidade do ego, consubstanciado pelo estabelecimento de um vínculo de apego em relações mais precoces positivas. Essas circunstâncias possibilitam "a tolerância de conteúdos mais arcaicos e a percepção de condições pessoais para reordená-los, emergindo da situação com uma resposta de melhor nível". Não obstante, se tais requisitos não são satisfeitos, "observa-se a fluencia de conteúdos do processo primário, com presença ou não de ansiedade e de recursos defensivos, mas não se verifica a reversibilidade do processo" (Cunha, Nunes, 1993, p.103). Conseqüentemente, o caráter patológico do processo "não é determinado pela profundidade da regressão, mas antes por sua natureza irreversível, pelo conflito que engendra e por sua interferência no processo de adaptação" (Arlow, Brenner, 1964, p.82).

Além de um afrouxamento de controles suscitado pela mobilização afetiva, Holt (1970) sugeriu outros motivos para os testes projetivos facilitarem a emergência da fantasia do sujeito. Fazendo referência mais específica ao teste de Rorschach, salientou que as próprias instruções favorecem a produção de imagens visuais, com desencorajamento implícito de idéias abstratas, que são "exclusivas do processo secundário" (Holt, 1970, p.272). Em segundo lugar, considerou que a ambigüidade dos estímulos não familiares e a abertura proporcionada pela natureza das instruções incentivam uma abordagem mais imaginativa e, em terceiro lugar, a suposta reputação do teste, como um instrumento de exploração de aspectos mais profundos da personalidade, desencadearia um estado de ansiedade que eliciaria conteúdos do processo primário.

Este tipo de regressão constitui, pois, uma função do ego. Se este é muito rígido, a capacidade de regressão pode ser obstaculizada ou limitada; quando é flexível, não só tolera modos de funcionamento mais primitivos como é capaz de controlar ou de interromper a regressão, promovendo o movimento inverso, reordenando, elaborando ou sintetizando o material da fantasia, de acordo com os requisitos do processo secundário e com as exigências implícitas na tarefa. Por outro lado, a perda do controle da regressão proporciona indícios patológicos, eventualmente emergentes mesmo num contexto de testagem neutro. Estes podem permitir identificar transtornos de pensamento, de valor diagnóstico em certos quadros nosológicos (Carr, Goldstein, 1981; Berg, 1983; Edell, 1987; Rosenberg, Miller, 1989; Gartner, Hurt, Gartner, 1989). Por outro lado, há muito vêm sendo feitas tentativas para criar um sistema de escore para a mensuração do material do processo primário (Klopfer et al., 1954; Holt, Havel, 1961; Holt, 1970). Já as vicissitudes do processo fornecem subsídios para o entendimento psicodinâmico, adquirindo significado numa perspectiva teórica.

Esta última alternativa, com sua ênfase no modelo clínico, coloca o psicólogo, "mais do que o teste, no centro do processo de avaliação", criando controvérsias com profissionais mais definidamente comprometidos com uma orientação psicométrica (Korchin, Schuldberg, 1981, p. 1147), além de ser criticada por sua associação com pressupostos psicanalíticos e vista ceticamente por grupos de profissionais adeptos de uma posição científica mais rigidamente sofisticada (Goldstein, Hersen, 1990).

Ora, considerando questões metodológicas, o enfoque projetivo está realmente em desvantagem na atualidade. Ainda que interpretação deste tipo deva ser feita dentro de "limites realísticos" (Exner, 1989, p.534), com base em evidências quantificáveis e não simplesmente buscando a significação isolada de indícios, como a proposta por "livros de receita", do tipo apresentado por Jolies (1992), muitas pesquisas devem ser desenvolvidas para esclarecer melhor a questão da projeção e equacioná-la mais cienti-

ticamente, antes que as técnicas projetivas mereçam o status que lhes foi atribuído em meados deste século e que se possa falar em medida projetiva sem pruridos polêmicos.

FORÇAS E FRAQUEZAS DAS TÉCNICAS PROJETIVAS

Pesquisas de opinião sobre técnicas projetivas e coletas de informações sobre o seu uso, ainda que nem sempre concordantes, ressaltam a sua importância.

Uma investigação entre 500 psicólogos da American Psychological Association (APA), cujos resultados foram divulgados em 1977, sugeriu que o status das técnicas projetivas não sofreu modificação desde um levantamento realizado em 1960. A grande maioria dos psicólogos utilizava técnicas projetivas, em baterias que incluíam também testes objetivos, embora reconhecesse suas limitações psicométricas (Wade, Baker, 1977). Não obstante, um estudo comparativo entre as opiniões de psicólogos clínicos de 103 departamentos acadêmicos (com programas aprovados pela APA), de 1968 e 1983, enfatizou o desenvolvimento de uma atitude negativa em relação às técnicas projetivas, especialmente entre os profissionais mais jovens. Estranhamente, embora desvalorizando tais instrumentos, ficou evidente a opinião concomitante de que deveriam ser oferecidos cursos sobre eles. Por outro lado, os autores (Pruitt, Smith, Thelen, Lubin, 1985) atribuíram a posição assumida pelos psicólogos mais jovens à sua falta de treinamento, porque as opiniões negativas se referiam inclusive a testes melhor considerados na comunidade científica, como o Rorschach que, conforme Blatt (1986, p.344), tem se caracterizado por "seu alto grau de refinamento e consolidação".

Outros levantamentos, revisados por Weiner (1986), feitos na década de 70 e início de 80, sempre registraram o uso e o ensino de técnicas projetivas. Todavia, esse autor analisa dados de 1983, de Lubin, Matarazzo e Larsen, que assinalaram uma modificação na ordem de preferência dessas técnicas, junto com instrumentos objetivos, que aparentemente parece sugerir uma diminuição de sua utilização. Porém, se forem considerados os percentuais, na realidade, tal diferença não é apreciável e há variação quanto aos "padrões de uso de testes em distintos ambientes clínicos" (Weiner, 1986, p.455).

De qualquer modo e por várias razões, a literatura específica chega a mencionar um declínio no interesse pela testagem clínica e na utilização de técnicas projetivas nos últimos anos, tanto como instrumentos de avaliação como na pesquisa, apesar de sua popularidade (Korchin, Schuldberg, 1981). Já um levantamento divulgado em 1985 chegou à conclusão de que "as principais técnicas projetivas do passado continuam sendo populares", não só entre os membros da Society for Personality

Assessment, mas entre os psicólogos clínicos em geral (Piotrowski, Sherry, Keller, 1985, p.117). Por outro lado, os resultados de urna pesquisa internacional mais recente sobre o uso de testes destacaram o TAT, o Rorschach e o M MPI como os instrumentos de personalidade mais populares (Hu, Oakland, 1991). Isto sugere que, agora como há cinqüenta anos atrás, conforme ponto de vista de trabalho anterior citado por Sargent (1945, p.282), "os clínicos não estão dispostos ou não são capazes de se basearem somente em critérios objetivos", o que significa que "continuam a suplementar métodos ideográficos de análise da personalidade com métodos nomotéticos", podendo-se concluir que, se são usados, é "porque preenchem uma função" (Mundy, 1972, p.795).

Entretanto, em se tratando de "medida projetiva", críticos poderiam afirmar, com Klein (1988, p.460), que "infelizmente, a popularidade dos testes não é um índice de excelência". Em verdade, se a referência a teste psicológico subentende uma definição tradicional mais estrita, poucos instrumentos preencheriam critérios mais aceitáveis (Berger, 1977) e a qualidade das técnicas projetivas seria fortemente questionada pela dificuldade de integrar seu tipo de abordagem com outras tradições científicas da Psicologia (Macfarlane, Tuddenham, 1976). Aliás, já em 1965, Zubin, Eron e Schumer salientavam o contraste entre "o método, o vocabulário e, mesmo, os requisitos dessa área" das técnicas projetivas com os da psicometria (Zubin, Eron, Schumer, 1965, p.2). Deste modo, é uma tarefa complexa tentar avaliar tais técnicas de acordo com as condições que, segundo Lanyon e Goodstein (1982), deveriam ser satisfeitas para que pudessem se caracterizar como instrumentos padronizados de avaliação.

A primeira condição seria a presença de estrutura da resposta. De forma muito diversa dos testes objetivos, os próprios estímulos envolvidos em técnicas projetivas são, em maior ou menor grau, carentes de estrutura, o que favorece "a liberdade de resposta" (Phares, 1984, p.283) e o encorajamento à fantasia, suscitando uma multiplicidade variegada de respostas, todas elas percebidas como material importante. Similarmente, a escassez de estrutura subentende algum grau de ambigüidade. Por outro lado, mesmo quando eventualmente os estímulos se apresentam com um pouco mais de estrutura e menos ambigüidade, como em algumas lâminas do TAT, são "as instruções que criam uma tarefa que é ambígua", de modo que o examinando desconhece os propósitos do examinador — aliás, característica "considerada como uma das forças mais importantes da técnica" (Martin, 1988, p.262) —, mas, assim, não é possível esperar que a resposta se caracterize por seu grau de estrutura, de qualquer maneira quantificável.

Na realidade, conforme ponto de vista de Cronbach, de sua edição de 1970, citado por Berger (1977, p.207), um teste "é um procedimento sistemático" que permite descrever o comportamento do sujeito, "seja através de uma escala numérica ou de um sistema de categorias". Então,

a questão da falta de estrutura da resposta em técnicas projetivas tem procurado ser equacionada com o desenvolvimento de sistemas de escore ou de categorização das respostas. Não obstante, a partir de uma definição mais tradicional de teste, tais esforços têm sido criticados por terem "mais o propósito de sumariar as respostas do que de quantificá-las" (Maloney, Ward, 1976, p.347). Por outro lado, mesmo alguns adeptos de técnicas projetivas têm certas restrições a esses recursos, porque tais sistemas não conseguem, a *priori,* ser suficientemente abrangentes, deixando de fora variáveis que poderiam levar a maior refinamento da análise do material clínico (Blatt, Berman, 1984). Entretanto, uma das críticas mais freqüentes é de que dificilmente há unanimidade entre os clínicos em relação ao uso dos sistemas de escore existentes (Vane, Guarnaccia, 1989). O teste de Rorschach, por exemplo, se caracterizou, durante muito tempo, pela convivência de cinco sistemas de escores, muitas vezes alterados individualmente pelos psicólogos, surgindo, ainda, sistemas "personalizados" (Exner e Exner, 1972, p.403). Tal situação, afortunadamente, tende a se normalizar, com o desenvolvimento do Sistema Compreensivo (Exner, 1978, 1980, 1983) que, num enfoque percepto-cognitivo, se concentra exatamente na estrutura da resposta (Erdberg, 1990).

Já no que se refere ao TAT, essa questão pode ser considerada mais séria em vista da diversidade das abordagens existentes, embora haja quem saliente a necessidade de "uma síntese dos vários métodos de escore e de interpretação" (Pollyson, Norris, Ott, 1985, p.28), em especial, como instrumento de pesquisa. Todavia, seu uso clínico "permanece mais informal e idiossincrásico" (Lanyon, Goodstein, 1982, p.60), coerente com sua tradição, que "é orientada para a interpretação impressionista e não para o escore formal" (Cronbach, 1990, p.622). Outros autores já acham que, se não existe um sistema padronizado de escore ou de categorização das respostas, não há garantia de que estas sejam analisadas com propriedade, porque a interpretação passaria a depender "da experiência, do talento e do julgamento do clínico, características que não podem ser ensinadas" (Vane, Guarnaccia, 1989, p. 11). Além disso, há outras restrições a essa abordagem, porque favoreceria a subjetividade, dando lugar a que "a mesma resposta possa ser interpretada de várias maneiras" (Martin, 1988, p.263) ou sob a ótica de diferentes referenciais teóricos.

Em relação a outras técnicas, são feitas críticas bastante semelhantes. No campo das técnicas projetivas gráficas, especialmente, exceto alguns Instrumentos para os quais foram desenvolvidos sistemas de escore, a interpretação freqüentemente é feita, recorrendo a exemplos fornecidos em manuais ou em catálogos de respostas qualitativas, como o de Jolies (1992). Mas o principal problema em relação à utilização desses recursos é que geralmente não são fornecidos dados de pesquisa para subsidiar a interpretação, notando-se que, às vezes, a hipótese interpretativa se baseia na observação de casos clínicos isolados (Cunha 1993b) ou em indícios pesquisados apenas em pacientes psiquiátricos (Hirt, Genshaft, 1976).

Alem desses aspectos discutidos, parece importante chamar a atenção para a quantidade de variáveis, muitas vezes de difícil controle ou pouco pesquisadas, que podem influenciar as respostas de crianças a técnicas projetivas, tornando ainda mais difícil os intentos de quantificá-las, como nível sócio-econômico, sexo e, naturalmente, idade, no desenho da figura humana (Cunha, 1992) ou, ainda, idade, QI e habilidade para a leitura, no Rorschach infantil (Klein, 1988). Desta maneira, se não existem subsídios adequados, o uso de técnicas projetivas com crianças é considerado com restrições (DiLorenzo, 1987) ou é definidamente contra-indicado (Klein, 1988).

Em face de tais considerações, difícil é a classificação de alguma técnica projetiva como um teste padronizado, se este for definido como aquele "em que os métodos de administração e escore são claros e objetivos" (Hamsher, 1990, p.256).

Outra condição seria a existência de normas. Mas a própria variabilidade e o número de respostas eliciadas, características das técnicas projetivas, que favorecem a projeção "de aspectos singulares e individuais da personalidade" (Maloney, Ward, 1976, p.346), dificultam o desenvolvimento de normas. Também a necessidade de administração individual, em muitos desses instrumentos, não facilita a coleta de dados normativos em grandes amostras (Vane, Guarnaccia, 1989). Apesar disso, o grupo de Exner vem suprindo o mercado com informações relevantes (Erdberg, 1990), além de enfatizar a importância de uma base empírica para as inferências clínicas (Exner, 1989). Todavia, no que se refere à maioria das técnicas projetivas, "os dados normativos ou inexistem completamente, são inadequados ou se baseiam em populações vagamente descritas" (DiLorenzo, 1987, p.75). Por outro lado, embora Mundy (1972) acredite que o caráter ambíguo dos estímulos favorece a individualização das respostas sem reflexos culturais, desde 1965, Zubin, Eron e Schumer salientaram a influência das variações culturais sobre a situação e os produtos da testagem. Assim, mesmo quando existem normas, é necessário previamente verificar a validade da técnica, quando utilizada em outras condições culturais (Lanyon, Goodstein, 1982). As restrições feitas ao uso de testes intelectuais, quando há diferenças culturais e subculturais, também repercutiram na área de técnicas projetivas (Cunha, 1993a). As várias versões do TAT, que vêm sendo desenvolvidas para "melhor entender as interações personalidade-cultura" (Jacquemin, 1982, p. 141), bem sugerem a necessidade de normas locais, assunto que vem sendo explorado por Constantino e seu grupo com a técnica TEMAS (Constantino, Malgady, Rogler, 1988), principalmente em seus esforços para padronizá-la para grupos minoritários (Constantino, Malgady, Rogler, 1988; Constantino et ai., 1992).

Além dos aspectos citados, Lanyon e Goodstein (1982, p.45) lembram que, embora seja possível registrar uma série de tentativas para

o desenvolvimento de normas para técnicas projetivas, não se pode afirmar que elas "tenham sido geralmente aceitas ou integradas na prática clínica", de modo que a interpretação "se torna complexa e demorada". Entretanto, há exceções e, indubitavelmente, o teste de Rorschach é uma delas, se levarmos em conta a acumulação de "dados normativos sobre uma variedade de grupos clinicamente relevantes" (Erdberg, 1990, p.397).

Já no que se refere à terceira condição, utilidade do instrumento, conforme Lanyon e Goodstein (1982, p.46), ela é determinada pelo "número de validades ou de relações válidas que um teste possui e pela importância ou significação dessas relações". Então, considerando, por exemplo, o teste de Rorschach, que é uma das técnicas que mais vêm sendo estudadas em termos da validação de hipóteses, Cronbach (1990) observa que os resultados de pesquisas ora se mostram marcantemente favoráveis, ora não. Não obstante, Erdberg (1990, p.389) chama a atenção para o fato de que certas diferenças entre os resultados se associam com o foco de conceptualização da natureza do teste. O foco percepto-cognitivo se concentra na estrutura das respostas do sujeito e, com este prisma, existe um grande acervo de "estudos de validade que vinculam variáveis estruturais do Rorschach ao comportamento alheio ao teste". Porém, quando o instrumento é utilizado como um estímulo à fantasia, o foco se dirige não para a estrutura das respostas, mas para as próprias verbalizações, sendo que, a partir de seu conteúdo, são levantadas hipóteses sobre a dinâmica interna. Neste caso, a questão da validade torna-se delicada e complexa. Assim, levando em conta as possíveis abordagens do teste, sua validade "é objeto de contínua controvérsia" (Lanyon, Goodstein, 1982, p.54). No entanto, esforços nesse sentido há muito vêm sendo feitos, destacando-se o trabalho de Zubin, Eron e Schumer (1965), que também pretenderam conceituai izar a técnica sob um referencial teórico diverso do enfoque junguiano original. Mais recentemente, registra-se o empenho de Weiner (1986), que vem trabalhando numa abordagem conceptual para estabelecer a validade de construto, embora salientando a importância também de um enfoque empírico, para documentar a base psicométrica das inferências de Rorschach sobre psicopatologia. Outros estudos vêm sendo desenvolvidos, mas não é possível tentar uma revisão mais completa dentro das dimensões deste capítulo.

Da mesma forma que o Rorschach, o TAT tem se mantido como uma técnica popularmente utilizada, dando origem a uma grande série de instrumentos congêneres, apesar da existência de forte controvérsia a respeito de sua validade, conforme revisões da literatura específica de Lanyon e Goodstein (1982) e de Lundy (1988). Aliás, este último autor demonstrou a complexidade do problema, ao tentar examinar as variáveis que podem estar em jogo quando "dois conjuntos de pesquisadores parecem obter diferentes resultados a partir do mesmo método" (Lundy, 1988, p.310), chegando à conclusão de que variações na situação de

testagem, em especial quanto às instruções, estariam afetando as respostas ao instrumento. Cronbach (1990) analisa, em seu livro, este e outros estudos, chamando a atenção para a importância de influências situacionais transitórias, que têm comprometido os resultados de pesquisas sobre validade. Aliás, Masling (1992) corrobora esses achados, mostrando o reflexo de variáveis situacionais e interpessoais, que acabam por ocasionar diferenças metodológicas, interferindo nos resultados de pesquisa e na situação de avaliação psicológica.

Ora, considerando a quantidade de pesquisas sobre o Rorschach e o TAT, bem como os aspectos controvertidos de seus resultados, tudo indica que o problema da validade das técnicas projetivas não é simples e parece não ter solução à vista, pelo menos em termos que possam ser satisfatórios para a tradição psicométrica mais ortodoxa. Não obstante, parecem estar emergindo novas tendências. Assim, na edição de 1956, Cronbach, citado por Masling (1992, p.616), salientava a posição galtoniana na pesquisa com testes, definindo teste como "uma amostra das respostas do sujeito a um estímulo padronizado impessoal". Já na edição de 1990, Cronbach (p.4) escreve que, embora seu texto se concentre mais em testes "padronizados", inclui, também, testes não padronizados, aos quais "se aplica a maioria dos conceitos usados para analisar" os primeiros. Discute variações nos métodos de colher informações, chamando a atenção, também, para diferenças na interpretação entre os aplicadores de testes, notando que "alguns têm como propósito 'medir' o indivíduo; outros, caracterizá-lo" (id. ibid., p.34). Tais propósitos definem estilos de testagem, numa abordagem psicométrica ou impressionista. Já chegando à questão da validação, salienta que os psicometristas confiam em interpretações com base quantificável a partir de normas estatísticas derivadas de estudos prévios, e não em "interpretações mais subjetivas, individualizadas". Ao contrário, "o impressionista está menos preocupado com validação formal", porque "validar 'retratos' é muito mais difícil do que validar predições numéricas. Com efeito, requer a validação do intérprete". Deste modo, "as abordagens psicométrica e impressionista diferem mais nitidamente na questão da confiança no psicólogo" (id. ibid., p.36).

Em resumo, no estado atual da arte, não parece possível transformar técnicas projetivas em psicométricas. Não obstante, "os psicólogos, hoje em dia, não apenas administram testes; eles realizam avaliações" (Goldstein, Hersen, 1990, p.3). Nesta perspectiva, num plano de avaliação, é possível utilizar os dois tipos de abordagem descritos por Cronbach, visando obter melhor consistência interna na seleção dos dados úteis e na integração dos resultados. Também, não é impossível que, com o incremento de pesquisas, seja viável emprestar melhores qualidades psicométricas a técnicas projetivas, especialmente pelo desenvolvimento de estudos de fidedignidade, validade e pelo estabelecimento de normas que permitam aumentar o grau de certeza nas inferências clínicas. Por

outro lado, além de uma perspectiva empírica, já vêm sendo feitos esforços no sentido de definir melhor a formulação teórica que associa as respostas ao teste com o comportamento na situação de vida para a qual são necessárias predições. Da mesma forma, também o referencial teórico que embasa a interpretação projetiva tem de se constituir um foco de novos estudos. Se não é possível testar uma teoria, construios teóricos podem ser investigados, desde que operacionalmente definidos (Cunha, Werlang, 1991a), dando renovado suporte a técnicas projetivas que provavelmente continuarão a ser utilizadas, sobreviventes que são de grandes controvérsias. Aliás, nota-se, também, que mesmo instrumentos vistos como medidas mais diretas e objetivas da personalidade, como o MMPI, vêm sendo interpretados em seus aspectos configuracionais, chegando a sínteses clínicas inspiradas na visão global da personalidade, uma das fortes características das técnicas projetivas.

Em vista de tudo isto, parece viável a previsão de Millón (1984, p.454) de um renascimento no campo da avaliação da personalidade, com integração de algumas formulações teóricas, advogando uma união instrumental em que condições empíricas e quantitativas se vinculem "com as qualidades clinicamente dinâmicas e integrativas, que caracterizam muitas técnicas intuitivas". Se esta é a perspectiva, só o futuro dirá.

REFERÊNCIAS BIBLIOGRÁFICAS

ABT, LE. Una teoria de la psicologia proyectiva. In: ABT, LE.. BELLAK, L. *Psicologia progrediva.* Buenos Aires: Paidós, 1%7. p.37-53.

ANZIEU. D. *Os métodos projetivos.* Rio de Janeiro: Campus. 1981.

ARLOW, J.A., BRENNER. C. *Psychoanalytic concepts and the structural theory.* New York: International Universities Press. 1964.

BELLAK, L Sobre los problemas del concepto de proyección: una teoria de la distorción aperceptiva. In: ABT, L.E.. BELLAK. L. *Psicologia proyectiva.* Buenos Aires: Paidós. 1967. p.25-36, 1967.

--------------. *El uso clinico de las pruebas psicológicas del TAT, CAT y SAT.* Mexico: Manual Moderno. 1979.

BELLAK, L. BELLAK, S. *Teste de apercepção temática (com figuras animais).* São Paulo: Mestre Jou, 1981.

BELLAK. L. HURVICH, M., GEDIMAN, H.K. *Ego functions in schizophrenics, neurotics and normals.* New York: Wiley-Interscience, 1973.

BERG, M. Borderline psychopathology as displayed by psychological tests. *Journal of Personality Assessment,* Portland (OR), v.47, n.2. p. 120-133,1983.

BERGER, M. Psychological testing. In: RUTTER, M., HERSON, L. *Child psychiatry,* modern approaches. London: Blackwell. 1977. p.306-333.

BERNSTEIN, J. Introducción: I. Métodos proyectivos. In: MURRAY. H.A. *Test de apercepción temática* (TAT). Buenos Aires: Paidós, 1951. p. 13-23.

BLATT, S.S. Whore have we been and where are we going? Reflections on 50 years of personality assessment *Journal of Personality Assessment,* Portland (OR), v.50, n.3, p.343-346 1986.

BLATT, S.S., BERMAN, W.H. A methodology for the use of the Rorschach in clinical research. *Journal of Personality Assessment.* Portland (OR), v.4ª, n.3, p.226-239, 1984.

CARR, L.C., GOLDSTEIN, E.G. Approaches to the diagnosis of borderline conditions by use of psychological tests. *Journal of Personality Assessment,* Portland (OR), v.45, n.6, p.563-574,1981.

CATTELL, R.B. Principios fundamentales de los tests de personalidad proyectivos o de percepción errónea *(misperception).* In: ANDERSON. H.H., ANDERSON. G.L. *Técnicas proyectivas de diagnóstico psicológico.* Madrid: Rialp, 1976. p.88-136.

CONSTANTINO. G. MALGADY, R.G., CASULLO, M.M.. CASTILLO. A. *La estandardización transcultural de TEMA en tres sub-culturas hispanas.* Madrid. 1992. Apresentado no Congreso Iberoamericano de Psicologia.

CONSTANTINO, G., MALGADY, R.G., ROGLER, L.H. *Technical manual:* the temas thematic apperception test. Los Angeles: Western Psychological Services, 1988.

CONSTANTINO, G, MALGADY. R.G., ROGLER. L.H.. TSUI, E.C. Discriminant analysis of clinical outpatients and public school children by TEMAS: a thematic apperception test for hispanic and blacks. *Journal of Personality Assessment.* Portland (OR), v.52, n.4, p.670-678, 1988.

CRONBACH, LJ. *Essentials of psychological testing.* 5.ed. New York: Harper and Row, 1990.

CUNHA. JA. Aspectos culturais dos testes psicológicos. *Psico.* Porto Alegre, v.24, n.l, p.69-74, 1993a.

------------- . *Psicodiagnóstico-R.* 4.ed.rev. Porto Alegre: Artes Médicas, 1993b. p.2I9-263: Bender.

------------- . *Problemas na avaliação de crianças na fase pré-escolar.* Simpósio : Diferencias socio-culturales y evaluación psicológica. Madrid, 1992. Apresentado no Congreso Iberoamericano de Psicologia.

CUNHA. J.A., WERLANG, B.G. *A estrutura edipica avaliada através do Teste das Fábulas.* Gramado (RS), 1991a. Apresentado no Primeiro Congresso Gaúcho de Psiquiatria.

. Respostas ao CAT-A em ISOpré-escolares. In: REUNIÃO ANUAL DA SOCIE DADE DE PSICOLOGIA DE RIBEIRÃO PRETO, 21.1991. *Anais...* Ribeirão Preto, 1991b.

CUNHA, J.A., NUNES, M.L.T., WERLANG, B.G. As respostas ao CAT-A na faixa pré-escolar. In: REUNIÃO ANUAL DA SOCIEDADE DE PSICOLOGIA DE RIBEIRÃO PRETO, 20, 1990. *Resumos...* Ribeirão Preto. 1990.

CUNHA, J.A.. NUNES, M.L.T., WERLANG, B.G., OLIVEIRA, M.S., WAGNER, A. CAT e a avaliação da qualidade projetiva das respostas. In: CUNHA, J.A. *Psicodiagnóstico-R.* 4.ed.rev. Porto Alegre: Artes Médicas. 1993. p.265-270.

DILORENZO, T. M. Standardized and projective tests. In: FRAME, C.L.. MATSON. J.L. *Handbook of assessment in childhood and psychopathology.* applied issues in differential and treatment evaluation. New York: Plenum, 1987. p.63-78.

EDELL, W.L. Role of structure in disordered thinking in borderline and schizophrenic disorders. *Journal of Personality Assessment.* Portland (OR), v.50, n.l, p.23-4I, 1987.

ERDBERG, P. Rorschach assessment In: GOLDSTEIN, G., HERSEN, M. (Eds.). *Handbook of psychological assessment* 2.ed. New York: Pergamon Press, 1990, p.387-399.

EXNER, J.E. Rorschach assessment. In: WEINER, J.B. (Ed.). *Clinical methods in psychology.* New York: Wiley, 1983. p.58-99.

---------------. Searching for projection in the Rorschach. *Journal of Personality Assessment,* Portland (OR), v.53, p.3, p.530-536, 1989.

---------------. *Sistema comprensivo del Rorschach.* 2,ed. Madrid: Pablo del Rio, 1980. v.1.

---------------. *Sistema comprensivo del Rorschach.* Madrid: Pablo del Rio, 1978. v.2.

EXNER, J.E., EXNER, D.E. How clinicians use the Rorschach. *Journal of Personality Assessment,* Portland (OR), v.36, p.403-408, 1972.

FERREIRA, A.B.H. *Novo dicionário da Lingua Portuguesa.* 2.ed. Rio de Janeiro: Nova Fronteira, 1986.

FREUD, A. *Infância normal e patológica.* Rio de Janeiro: Zahar, 1971.

FREUD, S. Formulações sobre os dois princípios de funcionamento mental. In: EDIÇÃO STANDARD brasileira das obras psicológicas completas de Sigmund Freud. Rio de Janeiro: Imago, 1969a. v.12. p.277-286.

---------------. Notas psicanalíticas sobre um relato autobiográfico de um caso de paranóia *(dementia paranoides).* In: EDIÇÃO STANDARD brasileira das obras psicológicas completas de Sigmund Freud. Rio de Janeiro: Imago, 1969b. v.12. p.23-115.

. O inconsciente. In: EDIÇÃO STANDARD brasileira das obras psicológicas comple tas de Sigmund Freud. Rio de Janeiro: Imago, 1974a. v. 14. p.185-245.

---------------. Totem e tabu. In: EDIÇÃO STANDARD brasileira das obras psicológicas completas de Sigmund Freud. Rio de Janeiro: Imago, 1974b. v.13. p.20-123.

---------------. Os chistes e sua relação com o inconsciente. In: EDIÇÃO STANDARD brasileira das obras psicológicas completas de Sigmund Freud. Rio de Janeiro: Imago, 1977. v.8.

---------------. A interpretação dos sonhos (segunda parte). In: EDIÇÃO STANDARD brasileira das obras psicológicas completas de Sigmund Freud. 2.ed. Rio de Janeiro: Imago, 1987a. v.5.

. As neuropsicoses de defesa. In: EDIÇÃO STANDARD brasileira das obras psico lógicas completas de Sigmund Freud. 2.ed. Rio de Janeiro: Imago, 1987b. v.3. p.51-72.

---------------. Observações adicionais sobre as neuropsicoses de defesa. In: EDIÇÃO STAN DARD brasileira das obras psicológicas completas de Sigmund Fred. Rio de Janeiro: Imago, 1987c. v.3. p.154-173.

---------------. Sobre os fundamentos para destacar da neurastenia urna síndrome específica, denominada "Neurose de Angústia". In: EDIÇÃO STANDARD brasileira das obras psico-lógicas completas de Sigmund Freud. 2.ed. Rio de Janeiro: Imago, 1987d. v.3. p.91-112.

GARTNER, J., HURT, S.W., GARTNER, A. Psychological signs of borderline personality disorder: a review of the empirical literature. *Journal of Personality Assessment,* Portland (OR), v.53, n.3, p.423-441,1989.

GOLDSTEIN, G, HERSEN, M. Historical perspectives. In: GOLDSTEIN, G., HERSEN, M. (Eds.). *Handbook of psychological assessment* 2.ed. New York: Pergamon Press, 1990. p.3-17.

HAMSHER, K. de S. Specialized neuropsychological assessment methods. In: GOLDSTEIN, G., HERSEN, M. (Eds.). *Handbook of psychological assessment.* 2.ed. New York: Pergamon Press, 1990. p.256-279.

HARTMANN, H. *Psicologia do ago e o problema da adaptação.* Rio de Janeiro: BUP. 1968.

HIRT, M., GENSHAFT, J.L Personality assessment and tests for children. In: SANKAR, S. (Ed.). *Mental health in children.* Westbury (NY): PJH, 1976. p.103-159.

HIRSCH, S.B., VERTHELYI, R., MENENDEZ DE RODRIGUEZ. F.M. *El Cat en el psicodiagnostico de niños.* Buenos Aires: Nueva Visión, 1979.

HOAR, M.W., FAUST, W.L. The Children's Apperception Test: puzzle and regular form. *Journal of Personality Assessment,* Portland (OR), v.37, p.244-247, 1962.

HOLT, R.R. Artistic creativity and Rorschach measures of adaptive regression. In: KLOPFER, E., MEYER, M.M., BRAWER, B. *Developments in the Rorschach technique.* New York: Harcourt Brace Jovanovich, 1970. v.3. p.263-320.

HOLT, R.R., HAVEL, J. A method for assessing primary and secondary process in the Rorschach. In: RICKERS-OVSIANKINA, MA. (Ed.). *Rorschach psychology.* New York: Wiley, 1961. p.263-315.

HU, S., OAKLAND, T Global and regional perspectives on testing children and youth: an empirical study. *International Journal of Psychoanalysis,* London, v.26, n.3, p.329-344,1991.

JACQUEMIN, A. Les variantes du "Thematic apperception test" pour l'étude des groupes culturels. *Revue Belge de Psychologie et de Pédagogie,* Brussels, v.44, n. 180, p. 135-144, 1982.

JOLLES, I. *A catalog for the qualitative interpretation of the House-Tree-Person* (H-T-P). Los Angeles: Western Psychological Services, 1992.

KLEIN, A. Questioning the clinical usefulness of projective psychological tests for children. In: CHESSS, S., THOMAS, A., HERZIG, M. *Annual progress in child psychiatry and child development.* New York: Brunner/Mazel. 1988.

KLOPFER, B., AINSWORTH, M.D., KLOPFER, W.G., HOLT, R.R. *Developments in the Rorschach technique.* New York: Harcourt. Brace and World, 1954. v. 1.

KORCHIN, S.J., SCHULDBERG, D. The future of clinical assessment. *American Psychologist,* Washington, D.C., v.36, n.10. p.1147-1158, 1981.

LANYON, R.I.. GOODSTEIN, L.D. *Personality assessment.* New York: Wiley. 1982.

LAPLANCHE, J., PONTALIS, J.B. *Diccionario de psicoanálisis.* 2.ed.rev. Barcelona: Labor, 1968.

LINDZEY, G. *Projective techniques and cross-cultural research.* New York: Appleton-Century-Crofts, 1961.

LUNDY, A. Instructional set and Thematic Apperception Test validity. *Journal of Personality Assessment.* Portland (OR), v.52, n.2. p.309-320, 1988.

MACFARLANE, J.W.. TUDDENHAM, R.D. Problemas planteados en la validación de las técnicas proyectivas. In: ANDERSON, H.H., ANDERSON. G.L. *Técnicas proyectivas de diagnóstico psicológico.* Madrid: Rialp, 1976. p.54-87.

MALONEY, M.P., WARD, *M.* P. *Psychological assessment* a conceptual approach. New York: Oxford Universities Press, 1976.

MARTIN. R.P. *Assessment of personality and behavior problems,* infancy through adolescence. New York: Guilford, 1988.

MASLING, J. The influence of situational and interpersonal variables in projective testing. *Journal of Personality Assessment,* Portland (OR), v.59, n.3, p.616-640. 1992.

MASSON, J.M. *Sigmund Freud Briefe am Wilhelm Fliess 1887-1904.* Frankfurt am Main: Fischer, 1986.

MILLÓN, T. On the renaissance of personality assessment and personality theory. *Journal of Personality Assessment.* Portland (OR), v.4ª, n.5, p.450-459, 1984.

MONTAGNA, M. E. *Análise e interpretação do CAT (teste de Apercepção Temática Infantil).* Sao Paulo: EPU, 1989.

MUNDY, J. The use of projective techniques with children. In: WOLMAN, B. (Ed.). *Manual of childpsychopathology.* New York: MacGraw-Hill, 1972. p.791-819.

MURRAY, H.A. *Test de Apercepción Temática (TAT).* Buenos Aires: Paidós, 1951.

MURSTEIN, B.I., PRYER, R.R. The Concept of projection: a review. *Psychological Bulletin,* Washington, D.C., v.56, n.5, p.353-374,1959.

PETERSON, C.A., SCHILLING, KM. Card pull in projective testing. *Journal of Personality Assessment.* Portland (OR), v.47, n.3. p.265-283. 1983.

PHARES, E.J. *Clinical psychology,* concepts, methods, and profession. Homewood (IL): Dorsey, 1984.

PIOTROWSKI, C, SHERRY. D.. KELLER, J.W. Psychodiagnostic test usage: a survey of the Society for Personality Assessment. *Journal of Personality Assessment.* Portland (OR), v.49, n.2, p.115-119, 1985.

POLLYSON, J.. NORRIS, D., OTT, E. The recent decline in TAT research. *Professional Psychology: Research and Practice.* Washington, D.C., v.16, n.2, p.26-28, 1985.

PRUITT. J.A., SMITH, M.C.. THELEN. M.H.. LUBIN, B. Attitudes of academic clinical psychologists toward projective technique: 1968-1983. *Professional Psychology: Research and Practice,* Washington, D.C., v. 16, n.6. p.781-788, 1985.

RAPAPORT, D. *El modelo psicanalitico, la teoria del pensamiento y las técnicas proyectivas.* Buenos Aires: Hormé, 1978.

. Sobre la teoria psicanalitica del pensar. In: GILL, M.M.. RAPAPORT, D. *Aportacio nes a la tècnica psicanalitica.* Mexico: Pax-Mexico, 1962, p.97-117.

-------------- . *Tests de diagnòstico psicològico.* Buenos Aires: Paidós, 1965.

RAPAPORT, D., GILL, M.M., SCHAFER, R. *Diagnostic psychological testing,* ed.rev. New York: International Universities Press, 1972.

RORSCHACH, H. *Psicodiagnóstico.* Buenos Aires: Médico-Quirúrgica, 1948.

ROSENBERG, P.H., MILLER, G.A. Comparing borderline definitions: DSM-III borderline and schizotypal personality disorders. *Journal of Abnormal Psychology.* Washington, D.C., V.98. n.2, p.161-169, 1989.

SARGENT, H. Projective methods: their origins, theory and application in personality research. *Psychological Bulletin.* Washington, D.C., v.42, n.5, p.257-293,1945.

SCHAFER, R. *Projective testing and psychoanalysis.* New York: International Universities Press, 1978.

---------------. *Psychoanalytical interpretation in Rorschach testing,* theory and application. New York: Grune and Stratton. 1954.

SHNEIDMAN, E. Projective Technique. In: WOLMAN, B. (Ed.). *Handbook of clinical psychology.* New York: McGraw-Hill, 1965. p.498-521.

VANE. J.R.. GUARNACCIA. V.C. Personality theory and personality assessment measures: How helpful to the clinician? *Journal of Clinical Psychology,* Brandon (VT), v.45, n. 1. p.5-19, 1989.

WADE. T.C., BAKER, T.B. Opinions and use of psychological tests - a survey of clinical psychologists. *American Psychologist,* Washington, D.C., p.874-882, Oct. 1977.

WEINER, I.B. Conceptual and empirical perspectives on the Rorschach assessment of psychopathology. *Journal of Personality Assessment,* Portland (OR), v.59, n.3. p.472-479, 1986.

------------- . The future of psychodiagnosis revisited. *Journal of Personality Assessment* Portland (OR). V.47, n.5. p.451-461. 1983.

WERLANG, B.G., CUNHA, J.A. Avaliação da personalidade sob o enfoque projetivo. In: CUNHA, J.A. *Psicodiagnóstico-R.* 4.ed.rev. Porto Alegre: Artes Médicas, 1993. p.123-129.

WOLFF, W., PRECKER, J.A. El movimiento expressivo y los métodos de la psicologia experimental profunda. In: ANDERSON. H.H.. ANDERSON, G.L. *Técnicas proyectivas de diagnóstico psicológico.* Madrid: Rialp. 1976, p.515-559.

ZUBIN, J.. ERON, L.D., SCHUMER, F. *An experimental approach to projective techniques.* New York: Wiley. 1965.

CAPÍTULO 14

A MEDIDA NA TEORIA DA AÇÃO RACIONAL

Maria Alice D'Amorim
Universidade Gama Filho

Segundo Fishbein (1980), é possível explicar e predizer o comportamento humano usando apenas alguns conceitos, ligados entre si por uma estrutura teórica de caráter compreensivo e global. Nesse texto, Fishbein atribui, pela primeira vez, o nome de *Teoria da Ação Racional* ao conjunto de seus trabalhos, justificando a escolha deste nome pelo pressuposto básico da teoria de que as pessoas agem de modo racional, processando implícita ou explicitamente as informações de que dispõem e utilizando-as como insumos em suas decisões. A teoria, porém, não assume que as informações processadas sejam necessariamente completas ou verídicas, apenas que, estando estas disponíveis, são usadas, e quando não o estão, ativamente buscadas. O autor afirma também nesse texto que a teoria está ainda incompleta e em busca de aperfeiçoamento, mas que a mesma tem-se revelado útil, pois fornece aos pesquisadores um enfoque sistemático para o estudo do comportamento volitivo em um grande número de áreas; além disto, a Teoria da Ação Racional pode servir como instrumento para a identificação de fatores que distinguem as pessoas que realizam determinado comportamento das que não o fazem, sejam estas diferenças explicadas pelo modelo teórico ou oriundas de variações sociodemográficas ou de personalidade.

O objetivo da Teoria da Ação Racional é compreender e predizer o comportamento, o que exige uma definição operacional deste comportamento. A teoria assume que a maioria das atividades humanas possui relevância social e está sob controle volitivo; como decorrência deste pressuposto, a teoria considera a intenção de realizar ou não uma atividade como o melhor preditor do comportamento futuro. Embora não exista uma correlação perfeita entre o dizer e o fazer, na maioria dos casos, salvo quando fatores situacionais interferem, a pessoa se comporta de acordo com a sua intenção. Como uma teoria deve ultrapassar o nível da predição dos fenômenos para chegar à sua explicação, torna-se necessário identificar os determinantes da intenção e, portanto, do comportamento.

Segundo a Teoria da Ação Racional, a intenção de realizar uma ação, chamada de intenção comportamental (I), está sob a influência de dois determinantes básicos, um pessoal e o outro social. O fator pessoal é a ponderação probabilistica das possíveis conseqüências da ação e sua avaliação em termos de nível de favorabilidade; este determinante leva o nome de atitude em relação ao comportamento (ATc). O fator social é a percepção, pela pessoa, das pressões sociais sofridas na realização ou não de um comportamento específico; como se trata da percepção das normas sociais, este fator é chamado de norma subjetiva (NS). De modo geral, a intenção de realizar uma atividade, tal como ir a uma festa, resulta de uma avaliação pessoal positiva das possíveis conseqüências desta ação e da percepção de que ela será socialmente aprovada. A importância relativa de cada um destes dois determinantes da intenção varia segundo a pessoa e o tipo de comportamento a ser realizado, tornando-se assim o cálculo do peso de cada um destes fatores uma questão empírica. O modelo teórico permitirá determinar, para cada ação estudada, o fator que mais contribuiu para a intenção comportamental. Esta relação entre as diversas variáveis do modelo pode ser expressa através de uma equação:

$$C \approx I = p_1 ATc + p_2 NS \; ,$$

onde C = comportamento, I = intenção comportamental, ATc = atitude em relação ao comportamento, NS = norma subjetiva ligada ao comportamento e p, e p_2 são os pesos empíricos de ATc e NS, respectivamente.

Assim, a intenção comportamental de uma pessoa seria o resultado da soma ponderada de sua atitude e norma subjetiva em relação ao comportamento estudado.

De acordo com a Teoria da Ação Racional, atitude e norma subjetiva são, ambas, influenciadas por crenças formadas com base na informação relevante disponível à pessoa. No caso da atitude, elas são chamadas de crenças comportamentais (CC) e estão ligadas às expectativas da pessoa de obter resultados favoráveis ou desfavoráveis com a execução do comportamento. As crenças mais facilmente evocadas pela pessoa em relação a um comportamento específico são chamadas de crenças modais salientes. A pessoa pondera cada uma das possíveis conseqüências de uma futura atividade, em termos de sua probabilidade de ocorrência; além disto, ela avalia estas conseqüências em uma escala de favorabilidade, (AC). O somatório dos produtos da probabilidade de ocorrência de cada uma das crenças por sua avaliação constitui uma medida indireta da atitude, conforme a equação:

$$ATc = CC_i \times AC_i \; ,$$

onde ATc = atitude com relação ao comportamento, CC_i = expectativa probabilistica de cada conseqüência e AC, = avaliação da favorabilidade de cada conseqüência.

Ao medir a força das crenças modais salientes e sua avaliação, obtém-se uma medida indireta da atitude que permite a sua predição, além de uma explicação acerca dos elementos que a compõem (Fishbein, Ajzen, 1975). A medida direta da atitude (ATc) é feita através de um diferencial semântico cujas dimensões são escolhidas de acordo com o comportamento estudado; em geral se usa bom-mau, útil—inútil, agradável-desagradável.

A norma subjetiva (NS) desenvolve-se através da percepção que o indivíduo tem de que pessoas significativas, chamadas referentes, aprovam ou desaprovam o comportamento que pretende realizar. Estas percepções são chamadas de crenças normativas (CN), dado o seu caráter social. Os referentes podem variar para uma mesma pessoa ou grupo de pessoas, conforme o comportamento estudado. A importância de cada referente depende do nível de motivação da pessoa para acatar a sua opinião; o somatório dos produtos da opinião de cada referente pela motivação para acatá-la constitui uma medida indireta da norma subjetiva, conforme a equação:

$$NS = CN_i \times M_i \ ,$$

onde NS - norma subjetiva ligada ao comportamento, CN_i = crenças acerca da opinião dos referentes e M_i - motivação para acatar as opiniões dos referentes.

A medida direta da norma subjetiva é feita através de uma única questão que engloba todas as pessoas cuja opinião é relevante para os indivíduos do grupo estudado, no caso do comportamento em questão, (Ajzen, Fishbein, 1972).

Correlações altas têm sido encontradas entre as medidas direta e indireta da atitude, bem como da norma subjetiva (Ajzen, Fishbein, 1980). Uma síntese do modelo da Teoria da Ação Racional pode ser vista na Figura 14-1, na página seguinte.

Segundo Fishbein (1980), o comportamento é explicado em nível global pela intenção; no nível seguinte, esta é explicada pela atitude e norma subjetiva; num terceiro nível, a atitude e a norma subjetiva são explicadas pelas crenças. Assim, em última análise, são as crenças de uma pessoa que explicam o seu comportamento. Como estas crenças representam a informação, correta ou incorreta, que a pessoa possui acerca do mundo, segue-se que o comportamento individual é determinado pela soma das informações que a pessoa possui. Deve-se porém observar que o aumento da compreensão das razões que levam ao comportamento não significa um melhor nível de previsibilidade, já que este pode ser predito apenas com o conhecimento da intenção. As crenças, atitude e norma subjetiva influenciam o comportamento indiretamente, através da intenção, não contribuindo, diretamente, para a sua
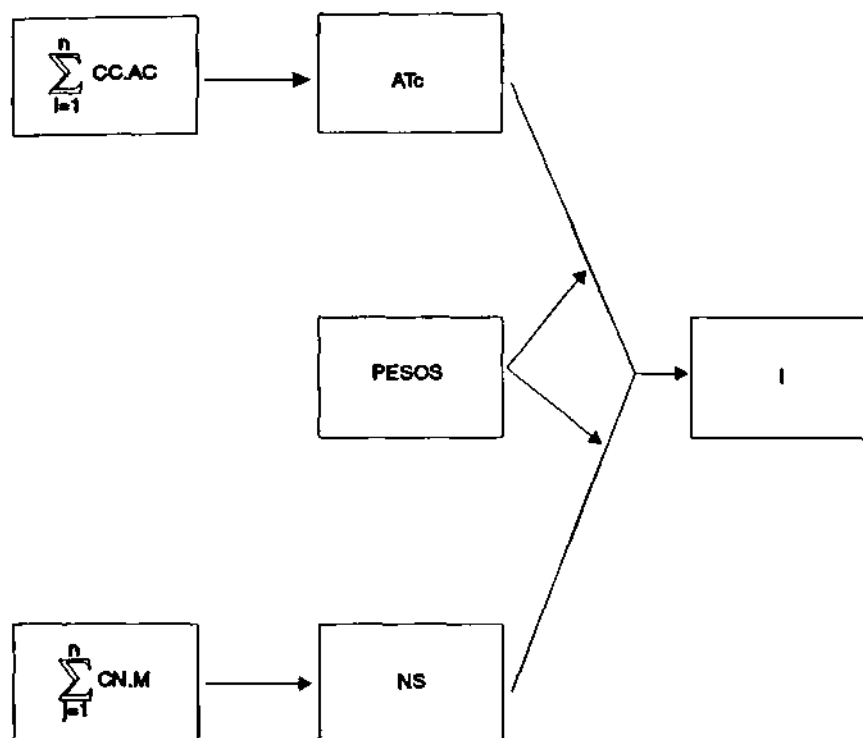
Fig. 14-1 — Elementos para predição e compreensão do comportamento segundo a Teoria da Ação Racional (adaptado de Fishbein, 1980, p.107).

predição. Como a medida direta do comportamento é muitas vezes difícil, a Teoria da Ação Racional preocupa-se especialmente com o estudo da intenção.

AFIRMAÇÕES BÁSICAS DA TEORIA DA AÇÃO RACIONAL

A Teoria da Ação Racional baseia-se em uma série de afirmações acerca das relações entre as diversas variáveis do modelo teórico.

1 — A Teoria da Ação Racional refere-se à predição e explicação de comportamentos específicos (Ajzen, Fishbein, 1973).

É importante distinguir os comportamentos específicos das categorias comportamentais, que são conjuntos de várias atividades e que não podem ser observadas como tais; assim, perder peso não é um comportamento e sim uma categoria comportamental, pois inclui atividades como não comer doces, fazer exercício, usar alimentos dietéticos, etc. Outro

conceito a ser isolado do comportamento é o de objetivo *(outcome);* passar no vestibular é um objetivo que exige do candidato uma série de comportamentos, como o de freqüentar um cursinho, estudar, fazer resumos ou cálculos, e ainda depende de fatores externos ao controle do candidato, tais como a dificuldade das provas e o número de vagas oferecidas.

A medida de ato único não pode prever o desempenho em categorias comportamentais; assim, observar que alguém toma café sem açúcar não garante que a pessoa esteja fazendo dieta para perder peso, já que pode tratar-se de uma preferência pessoal ou de um diabético. Para poder-se concluir que se trata de uma dieta visando perder peso, são necessários vários outros índices comportamentais.

2 — A obtenção de uma relação significativa entre as medidas de intenção e comportamento (declarar o voto e votar) depende do grau de correspondência existente entre estas medidas.

Todo critério comportamental envolve quatro aspectos: a **ação** em si, o **objetivo** para o qual esta se dirige, o **contexto** no qual é realizada e o **momento** em que ocorre. Cada um destes aspectos pode variar do geral ao particular, e a correspondência do nível de especificidade nas medidas de intenção e comportamento é indispensável para a obtenção de correlações significativas entre estas duas variáveis.

A categoria comportamental refere-se a um conjunto de ações e deve corresponder a uma intenção que inclua estas várias ações. Neste caso, deve-se construir um índice intencional que corresponda à categoria comportamental.

Assim, a atividade de comprar (ação) coca-cola (objetivo) no supermercado mais próximo de casa (contexto) no sábado pela manhã (momento) não é um índice válido da categoria comportamental de comprar bebidas, já que a sua especificidade não permite uma generalização. Ações específicas são previsíveis com base em intenções também específicas, como declarar a intenção de voto e votar. Não se pode, porém, assumir que intenções gerais ligadas a categorias comportamentais, corno perder peso, ou a objetivos, como passar no vestibular, venham a prever comportamentos específicos, tais como fazer ginástica ou matricular-se num cursinho.

3 — As medidas tradicionais de atitude em relação a um objeto estão sistematicamente ligadas a critérios de atos múltiplos, porém não apresentam correlações significativas com critérios de atos únicos (Ajzen, Fishbein, 1974).

A atitude de uma pessoa em relação a um objetivo não está ligada a um único comportamento realizado em vista de atingir este objetivo, mas sim a uma tendência geral a agir de modo a alcançá-lo. No critério de atos

múltiplos, cada comportamento descrito pode ser visto como um item de urna escala de atitude, podendo esta ser construída segundo a metodologia tradicional.

4 — O conhecimento das crenças comportamentais (CC) e de suas avaliações (AC), além da medida direta da atitude (AT), aumenta o nível de compreensão da intenção (I), mas não a sua previsibilidade.

A teoria considera a atitude (AT) como uma função das crenças comportamentais e de sua avaliação, sendo o valor destas últimas urna medida indireta da mesma. Neste caso, tanto a medida direta da atitude quanto a indireta seriam igualmente válidas, podendo tanto uma quanto a outra, ou ambas, estar correlacionadas com a intenção comportamental. Empiricamente, foram encontrados casos onde existe uma correlação fraca entre as duas medidas da atitude, com a medida indireta (as crenças) sendo melhor preditora da intenção. Na maioria destes casos, trata-se de uma melhor correspondência dos quatro aspectos dos critérios comportamentais — ação, objetivo, contexto e momento — entre as crenças comportamentais e a intenção, do que entre esta última e a medida direta da atitude. Assim, nas pesquisas de mercado, as crenças comportamentais e a intenção se referem diretamente à compra de um produto específico, enquanto a medida direta da atitude aborda uma preferência geral pelo produto, diferindo da medida da intenção no aspecto ação (comprar e preferir).

5 — O uso de crenças gerais ou pessoais influi na previsão da intenção. Fishbein e Ajzen (1974) distinguem as crenças gerais das pessoais; assim, a crença de que "fumar é prejudicial à saúde" é diferente da que afirma "ao fumar estou prejudicando a minha saúde", pois fumantes e não-fumantes concordariam com a primeira, porém muitos fumantes discordariam da segunda. Fishbein (1980) descreve uma pesquisa sobre o comportamento colocando as questões em nível geral e pessoal e utilizando crenças acerca de fumar e de não fumar. As diferenças entre estas quatro condições na previsibilidade da intenção comportamental podem ser vistas na Tabela 14-1.

Tabela 14-1 — Predições da intenção diferencial (I_F - I_{NF}) com base nas crenças, usando medidas gerais e pessoais

Crenças	Medida Geral	Medida Pessoal
Crenças acerca de fumar	0,518	0,553
Crenças acerca de não fumar	-0,445	-0,501
Diferença entre os dois tipos de crença	0,539	0,586

I_F - Intenção de fumar; I_{NF} - Intenção de não fumar
Adaptada de Fishbein (1980, p.95)

As crenças pessoais apresentam correlações positivas mais altas com a intenção diferencial (I_F - I_{NF}) do que as crenças gerais. O autor concluiu que, para melhor compreender e prever a intenção e o comportamento em áreas que implicam decisões importantes, deve-se considerar as crenças, a atitude e a intenção acerca de todas as possíveis alternativas e avaliar estas três variáveis no nível pessoal.

6 — As variáveis externas ao modelo, sejam elas demográficas ou de personalidade, atuam na intenção através das crenças, atitude e norma subjetiva, não podendo assim ser utilizadas na explicação básica do comportamento.

Segundo Fishbein (1980), uma pessoa não realiza um comportamento por ser altruísta ou egoísta, jovem ou velho, homem ou mulher, e sim porque ela crê que a sua *performance* lhe trará um maior número de conseqüências boas do que más e, também, porque percebe a aprovação de pessoas ou grupos em relação a este seu comportamento. As variáveis externas representam um quarto nível de explicação, isto é, explicam porque as pessoas têm determinadas crenças comportamentais e normativas e apresentam diferenças nestas crenças, ligadas às suas características pessoais. As variáveis externas explicam também as diferenças no nível de aceitação das pressões sociais.

A Figura 14-2, na página seguinte, apresenta o modelo da Teoria da Ação Racional, incluindo as variáveis externas. Como se pode ver, as variáveis externas exercem influência sobre as crenças comportamentais e suas avaliações, sobre as crenças normativas e a motivação para acatar as opiniões alheias. A Teoria da Ação Racional permite identificar o *locus* ou *loci* onde se exerce esta influência e explicar porque uma determinada variável externa está relacionada com o comportamento estudado, o que ajuda a esclarecer algumas das inconsistências da literatura atitude-comportamento.

A MEDIDA NA TEORIA DA AÇÃO RACIONAL

A medida de um ato único

A Teoria da Ação Racional utiliza, como instrumento de medida dos diversos elementos do modelo teórico, o questionário. A primeira decisão, de grande importância, é a escolha do comportamento específico a ser estudado na população alvo; como exemplo, poderíamos escolher a doação regular de sangue por estudantes universitários.

A construção do instrumento passa por diversas fases; na primeira, uma amostra de indivíduos, com as mesmas características dos futuros sujeitos (estudantes universitários), fornece, através de entrevistas, uma
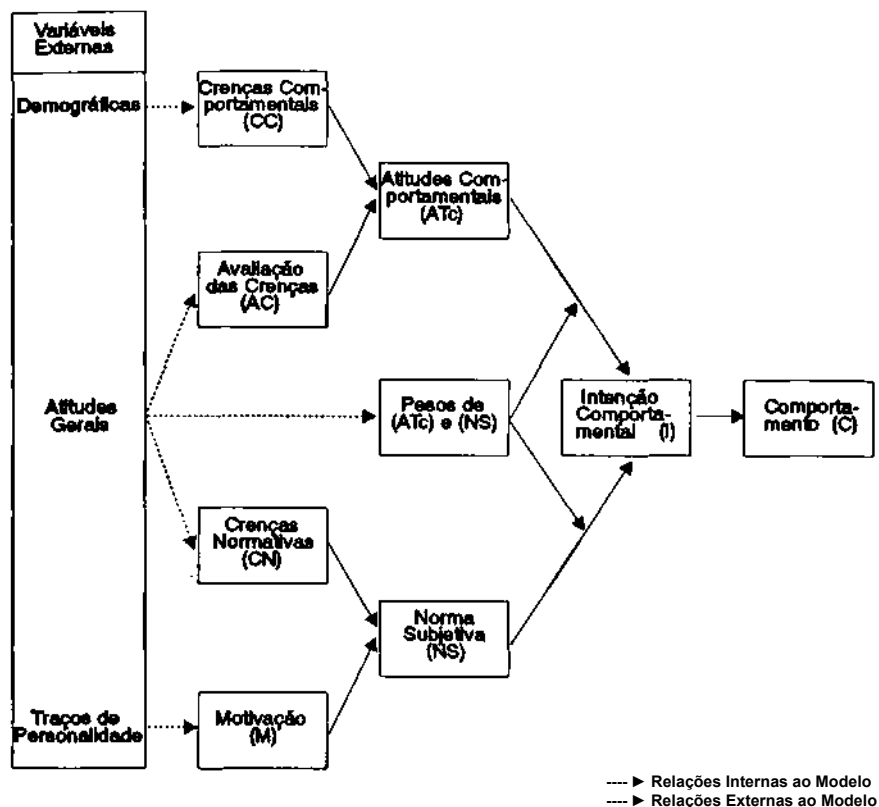
Figura 14-2 — Teoria da Ação Racional.
(Adaptado de Ajzen e Fishbein, 1980, p.8)

lista das vantagens e desvantagens de se tornar um doador regular de sangue. A pergunta é feita em caráter pessoal, não se tratando do comportamento geral de doar sangue, mas sim de uma ação do sujeito. As respostas irão constituir as crenças comportamentais (CC) do grupo, já que são uma listagem das possíveis conseqüências do comportamento em questão. Além desta pergunta, os sujeitos devem listar as pessoas cuja opinião acerca da sua doação de sangue seja importante para eles; estas pessoas, chamadas de referentes, embasam as crenças normativas do grupo acerca da aceitação social do comportamento estudado.

O número de sujeitos a serem entrevistados é determinado empiricamente, através do critério de, no mínimo, cinco entrevistas seguidas, onde as respostas obtidas nada acrescentem às vantagens e desvantagens do comportamento e ao número dos referentes.

Os dados obtidos sofrem uma análise de conteúdo que procura agrupar as respostas com o mesmo significado, dadas sob forma diferente; as mesmas vantagens ou desvantagens podem ser expressas de modo diverso. Após a análise, será calculada a freqüência de cada tipo de resposta, sendo as mais freqüentes consideradas como crenças modais salientes. Estas crenças vão constituir os itens do instrumento de pesquisa, sendo necessário um mínimo de aceitação por 75% dos membros do grupo entrevistado.

Numa segunda fase, o instrumento é elaborado sob a forma de questionário, utilizando escalas de 7 pontos que vão de provável (+3) a improvável (-3) para a maioria das variáveis: intenção (I), crenças comportamentais (CC) e normativas (CN), motivação para acatar a opinião dos referentes (M) e norma subjetiva (NS).

A avaliação das crenças é respondida numa escala de **bom** (+3) a **mau** (-3), e a atitude, através de um diferencial semântico cujas dimensões variam com o comportamento estudado. Um exemplo do instrumento é o que veremos a seguir.

Supondo que se estude o comportamento de doar sangue regularmente, com um grupo de universitários, e que, na primeira fase, tenham sido obtidas quatro crenças modais salientes e três referentes. As crenças obtidas são medidas duas vezes — quanto à probabilidade das possíveis conseqüências e quanto à sua avaliação. Os referentes também figuram duas vezes no questionário — como fontes de opinião e como objeto da motivação para acatar esta opinião.

As questões do instrumentos obedecem a uma ordem determinada de apresentação.

Atitude (AT) — Medida direta — um único item constituído pela soma das respostas obtidas nas várias dimensões de um diferencial semântico.

1 — O fato de tornar-me um doador regular de sangue é:

útil -- inútil
+3 +2 +1 0 -1 -2 -3

bom -- mau
+3 +2 +1 0 -1 -2 -3

agradável -- desagradável
+3 +2 +1 0 -1 -2 -3

Crenças comportamentais (CC) — as que foram consideradas crenças modais salientes na primeira fase.

O fato de tornar-me um doador regular de sangue:

2 — Me fará correr o risco de pegar AIDS.

provável -- improvável

+3 +2 +1 0 -1 -2 -3

3 — Ajudará as pessoas que precisam de transfusões.

provável -- improvável

+3 +2+1 0 -1 -2 -3

4 — Fará com que eu fique fraco e acabe doente.

provável -- improvável

+3 +2 +1 0 -1 -2 -3

5 — Me fará correr risco de vida, se sofrer um acidente.

provável -- improvável

+3 +2+1 0 -1 -2 -3

Avaliação das crenças (AC).

O fato de tornar-me um doador de sangue:

6 — Me fará correr o risco de pegar AIDS.

bom -- mau

+3 +2 +1 0 -1 -2 -3

7 — Ajudará as pessoas que precisam de transfusões.

bom -- mau

+3 +2 +1 0 -1 -2 -3

8 — Fará com que eu fique fraco e acabe doente.

bom -- mau

+3 +2+1 0 -1-2 -3

9 — Me fará correr risco de vida, se sofrer um acidente.

bom -- mau

+3 +2 +1 0 -1 -2 -3

Norma subjetiva (NS) — Medida direta.

10 — As pessoas cuja opinião é importante para mim acham que eu devo tornar-me um doador regular de sangue.

provável -- improvável
+3 +2 +1 0 -1 -2 -3

11 — Meu médico acha que eu devo tornar-me um doador regular de sangue.

provável -- improvável
+3 +2 +1 0 -1 -2 -3

12 — Meu pai acha que eu devo tornar-me um doador regular de sangue.

provável -- improvável
+3 +2 +1 0 -1 -2 -3

13 — Meu (minha) companheiro(a) acha que eu devo tornar-me um doador regular de sangue.

provável -- improvável

+3 +2 +1 0 -1 -2 -3

Motivação para concordar com os referentes (M).

14 — Eu costumo acatar as opiniões do meu médico.

provável -- improvável

+3 +2 +1 0 -1 -2 -3

15 — Eu costumo acatar as opiniões do meu pai.

provável -- improvável
+3 +2 +1 0 -1 -2 -3

16 — Eu costumo acatar as opiniões do(a) meu (minha) companheiro(a).

provável-- improvável

+3 +2 +1 0 -1 -2 -3

17 — Eu pretendo tornar-me um doador regular de sangue.

provável-- improvável
+3 +2 +1 0 -1 -2 -3

A medida em situações de escolha

Em muitos casos estamos diante de uma escolha entre dois ou mais comportamentos mutuamente exclusivos. Nestes casos, a medida da atitude deve ser feita em relação a cada uma das alternativas, para melhorar o nível de previsão da intenção comportamental e, através dela, a previsão do comportamento. Assim, se o indivíduo tem de escolher entre dois comportamentos, como fazer ou não dieta para perder peso, a diferença algébrica entre os escores da atitude em relação a cada uma das alternativas será melhor preditora da intenção que o escore atitudinal em relação a qualquer um dos dois comportamentos.

No caso de várias alternativas, como beber refrigerante, suco de fruta ou cerveja, a ordenação pelos sujeitos destas alternativas será a melhor preditora de sua intenção de beber cerveja, já que esta terá o primeiro lugar na sua ordenação. No caso da norma subjetiva, a opinião dos referentes e a motivação para acatá-la também devem ser medidas para todas as alternativas, a fim de melhorar a predição da intenção. Estas medidas devem manter para cada alternativa a unidade de ação (beber), contexto (local) e momento (dia, hora), embora variem os objetivos (bebidas).

Um exemplo da medida em situação de várias alternativas foi dado por Ajzen e Fishbein (1969), quanto às possíveis atividades de 100 estudantes universitários nas noites de sexta-feira.

Os autores listaram oito possíveis atividades, mutuamente exclusivas, que poderiam ser realizadas pelos estudantes:

1 — Ira uma festa
2 — Visitar uma exposição de arte moderna
3 — Ver um filme na televisão
4 — Ira um concerto
5 — Jogar póquer
6 — Irão cinema
7 — Participar de um grupo de debates
8 — Ler um livro policial

Medidas usadas no estudo

Atitude em relação ao objeto da atividade (AT) — Para cada uma das atividades, os sujeitos deveriam dar a sua avaliação em um diferencial semântico com quatro dimensões: bom-mau, agradável-desagradável, interessante-desinteressante e compensador-punitivo, em escalas de sete pontos. Os escores dos sujeitos foram calculados somando as suas respostas a cada uma das quatro dimensões, para cada uma das atividades. Total de oito medidas.

Atitude em relação ao comportamento (ATc) — Os mesmos dados do diferencial semântico foram somados para cada atividade, totalizando separadamente os resultados de cada dimensão do diferencial semântico. Assim, "ir a uma festa" era avaliada em cada dimensão do diferencial, sendo a soma dos dados considerada um índice da atitude em relação à atividade. Total de oito medidas.

Norma subjetiva pessoal (NSp) — Obtida para cada uma das oito atividades, com um único item totalizando oito medidas.

Eu, pessoalmente, acho que devo ir a uma festa na noite de sexta.

Provável _ _ _ _ _ _ _ _ _ Improvável

Norma subjetiva social (NSs) — Percepção da opinião dos amigos; um item para cada atividade. Total de oito medidas.

Meus amigos acham que eu devo ir a uma festa na noite de sexta.

Provável _ _ _ _ _ _ _ _ _ Improvável

As intenções comportamentais foram medidas de três maneiras:

Intenções específicas acerca de cada atividade (I) — medidas através de uma escala para cada uma das atividades, constituindo, assim, uma medida de ato único. Total de oito medidas.

Numa noite de sexta-feira

Eu iria _ _ _ _ _ _ _ _ _ Eu não iria
a uma festa

Intenções diferenciais (Id) — isto é, a escolha entre duas alternativas; as oito atividades foram combinadas em 28 pares apresentados em ordem randômica; os autores usaram aqui uma medida em situação de escolha diàdica, com um total de 28 medidas.

Numa noite de sexta-feira

Eu iria ao cinema - - - - - - - - - e u iria a um concerto

Intenções ordinais (Io) — medidas através da ordenação das oito atividades, de acordo com as preferências dos sujeitos. Total de oito medidas.

Com base nestas medidas, dois tipos de índices foram computados:

1 — Diferenças algébricas para cada um dos 28 pares de atividades para as medidas (AT — ATc), NSp e NSs. Estes índices foram chamados pela autora de ATd, NSpd e NSsd.

2 — A ordem de preferência de cada um dos oitos ATd, NSpd e NSsd para cada um dos 100 sujeitos. Estes índices receberam os nomes de ATo, NSpo e NSso.

0 questionário incluiu, assim, seis conjuntos de medidas para cada uma das oito atividades: AT, ATc, NSp, NSs, I e lo, num total de 4ª; além disto, incluía 28 pares de comparações, elevando o total das medidas a 76.

Resultados

Uma análise de variância foi usada para cada uma das 76 medidas.

1 — A predição da intenção comportamental **de ato único** (I) foi feita através de uma regressão múltipla, tendo como variáveis independentes ATc, NSp e NSs. Os resultados podem ser vistos na tabela 14-2.

Tabela 14-2 — Correlações múltiplas de I com ATc, NSp e NSs (N=100)

Comportamentos	AT C-I		N .p-I		N Ss-I		R
	R	B	R	B	R	B	
Ir a uma festa	0,523	0,077	0,815	0,714	0,587	0,083	0,819
Ir a uma exposição	0,670	0,440	0,630	0,272	0,437	0,128	0,724
Ver filme na TV	0,567	0,225	0,662	0,423	0,439	0,180	0,709
Ir ao concerto	0,665	0,303	0,713	0,376	0,598	0,249	0,787
Jogar póquer	0,668	0,227	0,767	0,502	0,591	0,158	0,794
Ir ao cinema	0,640	0,190	0,782	0,648	0,499	0,190	0,794
Participar de um debare	0,669	0,252	0,702	0,335	0,678	0,300	0,779
Ler livro policial	0,538	0,292	0,543	0,268	0,513	0,323	0,684

Todas as correlações são significativas a $p < 0,01$.

Os coeficientes múltiplos foram altos, mostrando o poder do modelo de Fishbein e Ajzen para prever as intenções comportamentais. Os maiores pesos vêm da variável NSp, norma subjetiva pessoal, dando suporte à afirmação teórica de que as crenças normativas são auxiliares importantes da atitude na previsão da intenção. Todos os valores beta para a norma subjetiva pessoal foram significativos a $p < 0,05$, indicando sua contribuição na previsão da intenção.

A medida de atitude em relação ao objeto da atividade (AT) apresentou correlações significativas de $p< 0,01$ com a atitude em relação ao comportamento (ATc) e às normas subjetivas pessoal (NSp) e social (NSs), bem como com a intenção (I).

2 — A predição da intenção comportamental **diferencial** (Id) foi feita inicialmente comparando as correlações entre a intenção diferencial e a atitude diferencial (ATd) com as atitudes em relação a cada um dos comportamentos. As correlações entre as duas medidas diferenciais, intenção e atitude, foram, em geral, mais altas que aquelas entre a intenção diferencial (Id) e as medidas de atitude em relação às várias atividades; as únicas exceções foram a atitude em relação à festa quando comparada à

leitura de uma estória policial ou a ver um filme na televisão. Ir a uma festa foi muito bem avaliada pelos sujeitos e, nestas duas comparações, a atitude em relação à atividade foi suficiente para uma melhor previsão da intenção diferencial. No caso da norma subjetiva, a mesma comparação foi feita entre as medidas ligadas a cada atividade e os índices NSpd e NSsd, quanto à sua capacidade de prever a intenção diferencial (Id) para cada atividade. Em geral, os resultados foram semelhantes aos obtidos com as atitudes. Em apenas quatro casos, o índice diferencial pessoal (NSp) apresentou previsões da intenção superior ao valor diferencial (NSpd), e dois deles envolviam a atividade de ir a uma festa. Para a norma subjetiva social (NSs), todos os índices diferenciais (NSsd) apresentaram correlações superiores com a intenção diferencial (Id).

Os resultados da regressão múltipla, tendo como variável dependente a intenção diferencial (Id) e como variáveis independentes ATd, NSpd e NSsd, aparecem na Tabela 14-3.

Tabela 14-3 — Correlações múltiplas de Id com ATd, NSpd e NSsd (N=100)

Variáveis	ATc -Id		NSpd-Id		NSsd-Id		R
	R	B	R	B	R	B	
01 Cinema-concerto	0,587	0,129	0,677	0,339	0,677	0,374	0,747
02 Festa-pôquer	0,436	-0,029	0,624	0,522	0,528	0,172	0,635
03 Livro-debate	0,727	0,378	0,378	0,714	0,341	0,195	0,800
04 TV-concerto	0,653	0,239	0,609	0,234	0,626	0,307	0,727
05 Livro-exposição	0,676	0,381	0,619	0,189	0,622	0,300	0,749
06 Festa-concerto	0,487	0,075	0,616	0,326	0,629	0,410	0,704
07 Exposição-cinema	0,638	0,323	0,622	0,227	0,998	0,288	0,721
08 Pôquer-concerto	0,691	0,202	0,716	0,264	0,751	0,451	0,818
09 Cinema-livro	0,622	0,332	0,563	0,173	0,614	0,356	0,773
10 Debate-festa	0,544	0,114	0,640	0,292	0,668	0,400	0,730
11 Livro-TV	0,437	0,099	0,562	0,371	0,475	0,200	0,589
12 Livro-festa	0,653	0,386	0,596	0,304	0,447	0,209	0,726
13 Exposição-concerto	0,518	0,277	0,518	0,250	0,482	0,212	0,612
14 Debate-concerto	0,616	0,219	0,643	0,230	0,704	0,437	0,769
15 Festa-livro	0,337	0,167	0,627	0,4%	0,501	0,231	0,657
16 Pôquer-TV	0,645	0,170	0,749	0,504	0,601	0,216	0,786
17 Pôquer-livro	0,593	0,167	0,664	0,391	0345	0,327	0,743
18 Cinema-TV	0,714	0,320	0,709	0325	0,649	0,357	0,791
19 Livro-concerto	0,584	0,256	0,586	0,269	0,576	0,306	0,6%
20 Exposição-festa	0,434	0,129	0,548	0,218	0,617	0,452	0,674
21 Debate-pôquer	0,738	0318	0,743	0,339	0,710	0,261	0,818
22 Cinema-festa	0,494	0,071	0,632	0,413	0360	0,293	0,678
23 Debate-cinema	0,741	0,438	0,689	0,209	0,681	0,245	0,795
24 Pôquer-exposição	0,747	0,285	0,744	0338	0,686	0,297	0,817
25 Debate-TV	0,770	0,446	0,701	0,228	0,685	0,262	0,826
26 Póquer-cinema	0,715	0,221	0,775	0,442	0,680	0,217	0,805
27 Debate-exposiçáo	0,611	0,151	0,667	0330	0,704	0,429	0,784
28 TV-exposição	0,792	0,559	0,687	0,183	0383	0,159	0,813

Todas as correlações são significativas a $p < 0,01$.

Como pode ser visto, as três variáveis independentes foram relevantes e estão altamente correlacionadas com a intenção comportamental diferencial. A importância relativa de cada preditor variou segundo as atividades comparadas. Na maioria dos casos, a norma subjetiva pessoal (NSpd) teve maior peso na previsão da intenção que as atitudes e as opiniões dos amigos (NSsd), o que fica claro com o exame dos valores beta.

As correlações múltiplas entre as três variáveis independentes foram altas, fornecendo evidência da utilidade do modelo da Teoria da Ação Racional e de seu poder preditivo em situações de escolha.

3 — Os sujeitos ordenaram as oito atividades segundo a probabilidade de realizá-las numa noite de sexta-feira. A ordenação destas intenções para cada atividade (Io) foi correlacionada com a atitude em relação à atividade (ATc), com as normas subjetivas pessoal (NSp) e social (NSs), específicas a cada comportamento, bem como com os índices de ordenação da atitude e norma subjetiva (ATo, NSpo, NSso). Previu-se que estes últimos teriam, com a intenção, correlações mais altas que as medidas simples (ATo, NSp e NSs). Os resultados das correlações entre Io, ATc, NSp, ATo, NSpo e NSso podem ser vistos na Tabela 14-4.

Tabela 14-4 — Correlações entre Io e ATc,
NSp, NSs, ATo, NSpo e NSso (N=100)

Fatores	Io Festa	Io Exposição	Io TV	Io Concerto	Io Póquer	Io Cinema	Io Debate	Io Livro
ATc	0,539	0,616	0,615	0,603	0,607	0,567	0,616	0,429
NSp	0,698	0,623	0,694	0,670	0,739	0,657	0,589	0,418
NSs	0,623	0,502	0,455	0,603	0,630	0,496	0,664	0,372
ATo	0,420	0,649	0,674	0,720	0,753	0,640	0,765	0,648
NSpo	0,579	0,713	0,752	0,752	0,845	0,775	0,664	0,555
NSso	0,612	0,622	0,541	0,713	0,758	0,624	0,675	0,520

Todas as correlações são significativas a $p < 0,01$.

Os dados obtidos suportaram a hipótese. As predições baseadas nos índices de ordenação foram melhores, tendo como base correlações mais altas. A exceção apareceu com a atividade "ir a uma festa". Os autores verificaram ainda o peso das variáveis independentes ATo, NSpo e NSso sobre a variável dependente Io. Os resultados aparecem na Tabela 14-5.

Pode-se notar que todas as correlações encontradas foram significativas, sendo especialmente altas aquelas entre a norma subjetiva pessoal (NSpo) e a intenção (Io), como pode ser constatado pelos valores beta

Tabela 14-5 — Correlações entre Io e ATo, NSpo e NSso (N=100)

Comportamentos	AT)-lo		NSp o-lo		NSs o-lo		R
	R	B	R	B	R	B	
Ir a uma festa	0,420	0,074	0,579	0,299	0,621	0,434	0,688
Ir a uma exposição	0,649	0,209	0,713	0,434	0,622	0,276	0,785
Ver filme na TV	0,647	0,160	0,788	0,603	0,541	0,136	0,810
Ir ao concerto	0,720	0,318	0,752	0,350	0,713	0,317	0,847
Jogar póquer	0,735	0,171	0,845	0,530	0,758	0,287	0,891
Ir ao cinema	0,646	0,135	0,775	0,606	0,624	0,101	0,787
Participar de um debate	0,765	0,487	0,664	0,236	0,654	0,213	0,818
Ler livro policial	0,648	0,373	0,555	0,287	0,520	0,239	0,725

Todas as correlações são significativas a $p < 0,01$.

A medida de atos únicos e de atos múltiplos

Fishbein e Ajzen (1974) realizaram um estudo elaborando medidas de ato único através de técnicas tradicionais de construção de escalas de atitudes e também critérios de ato único e de ato múltiplo. Os autores utilizaram 100 itens que descreviam comportamentos religiosos do tipo: rezar antes das refeições, fazer cursos de religião ou doar dinheiro a instituições de caridade. Inicialmente, estes itens foram avaliados por 62 sujeitos em termos de relatos de comportamento passado e por outros 63 em termos de intenção comportamental.

Os 100 itens serviram de base aos autores para elaborarem seis escalas:

1) — duas escalas do tipo Guttman foram construídas, uma para os auto-relatos e outra para as intenções, com oito itens cada. Os coeficientes de reprodutibilidade foram respectivamente 0,897 e 0,919, e os coeficientes marginais mínimos foram de 0,701 e 0,709;

2) — duas escalas tipo Likert foram elaboradas para os auto-relatos e intenções, com 20 itens cada, sendo usados os itens cuja correlação com os escores globais era mais alta. A variação das correlações item-total variou de 0,37 a 0,66 para os auto-relatos e de 0,43 a 0,70 para as intenções;

3) — a metodologia de Thurstone foi usada na construção de duas escalas, com 13 itens cada.

Além das seis escalas, os autores estabeleceram um critério de ato múltiplo, através de um grupo independente de 37 sujeitos que avaliaram os 100 itens, julgando-os numa escola de 11 pontos, cada um deles representando um nível de atitude, de favorável a desfavorável à Igreja. Os itens receberam valores positivos ou negativos, de acordo com esta classificação.

Para cada um dos 125 sujeitos iniciais, o auto-relato ou a intenção de realizar uma atividade positiva ou de não realizar uma negativa recebeu o valor um; os demais casos receberam o valor zero. O critério de ato múltiplo foi elaborado somando os pontos obtidos nos 100 itens para o auto-relato de comportamento (62 sujeitos) e para as intenções (63 sujeitos).

O critério de ato único foi obtido calculando a média das correlações de cada um dos 100 itens para o auto-relato e para a intenção. A Tabela 14-6 apresenta as correlações obtidas entre os três tipos de escala e os dois tipos de critério, para auto-relatos e intenções.

Tabela 14-6 — Correlações entre escalas comportamentais e critérios comportamentais

MEDIDAS	AUTO-RELATOS			INTENÇÕES		
Escalas Comportamentais	G	L	T	G	L	T
Gutman (G) Likert (L) Thurstone (T) Critério de atos múltiplos Critério de ato único	0,776 0,631 0,699 0,143	0,792 0,898 0,178	0,789 0,156	0,851 0,553 0,750 0,189	0,793 0,915 0,229	0,797 0,202

$r_{0,05}$=0,250; $r_{0,01}$=0.325 Adaptado de Fishbein (1974, p.64)

A observação da tabela mostra que as escalas apresentam, entre si, correlações altas assim como com o critério de ato múltiplo, o mesmo não acontecendo no caso do critério de ato único, baseado em comportamentos específicos.

CONCLUSÃO

A exposição sumária da Teoria da Ação Racional, bem como das suas aplicações a diversos tipos de comportamento, demonstra claramente a sua utilidade. O valor heurístico do modelo tem sido confirmado pela massa de pesquisas que originou desde a sua elaboração.

Críticas teóricas e metodológicas têm sido publicadas, porém o número de estudos que utilizam o modelo tem aumentado. Como foi dito inicialmente, a teoria está em desenvolvimento e várias sugestões interessantes já foram feitas, inclusive a de um dos autores, que acrescentou às variáveis atitude e norma subjetiva um terceiro determinante da intenção, a percepção do próprio controle sobre o comportamento em questão (Ajzen, 1988), passando a teoria a denominar-se Teoria da Ação Planejada.

REFERÊNCIAS BIBLIOGRÁFICAS

AJZEN, I. *Attitudes, personality and behavior.* [S.I.]: Milton Keynes Open University Press, 1988.

AJZEN, I., FISHBEIN, M. Attitudes and normative beliefs as factors influencing behavioral intentions. *Journal of Personality and Social Psychology,* Washington, D.C.,v.21,n. 1, p. 1-9, 1972.

--------------. Attitudes towards objects as predictors of single and multiple behavioral criteria. *Psychological Review,* Washington, D.C., v.81, n.l, p.59-74, 1974.

--------------. Attitudinal and normative variables as predictors of specific behaviors. *Journal of Personality and Social Psychology,* Washington, D.C., v.27, n.l, p.41-57,1973.

------------- . The prediction of behavioral intentions in a choice situation. *Journal of Experimental Social Psychology,* New York, v.5, p.400-416, 1969.

--------------. *Understanding attitudes and predicting social behavior.* Englewood Cliffs (NJ): Prentice-Hall. 1980.

FISHBEIN, M. A theory of reasoned action: some applications and implications. In: HOWE, H.E., PAGE, M.M. (Eds.). *Beliefs, attitudes and values:* Nebraska Symposium on Motivation, 1979. Lincoln (NB): University of Nebraska Press, 1980.

FISHBEIN, M., AJZEN, I. *Beliefs, attitude and intention:* an introduction to theory and research. Reading (MA): Wesley, 1975.

CAPITULO 15

DESENVOLVIMENTO DE INSTRUMENTO PARA LEVANTAMENTO DE DADOS *(SURVEY)*

Hartmut Günther
Instituto de Psicologia
Universidade de Brasilia

A expressão 'levantamento de dados', tradução do inglês *survey, é* definida por Fink e Kosecoff (1985, p.13) como um "método para coletar informação de pessoas acerca das suas idéias, sentimentos, planos, crenças, bem como origem social, educacional e financeira". Implícita nesta definição é sua aplicabilidade às mais diversas áreas das ciências sociais. O instrumento para este tipo de coleta de dados é o questionário, que pode ser administrado tanto em interação pessoal — em forma de entrevista individual ou por via telefônica — quanto auto-aplicado — seja após envio via correio ou em grupos. A Figura 15-1 mostra os componentes do processo de levantamento de dados, bem como uma primeira orientação para o presente capítulo.
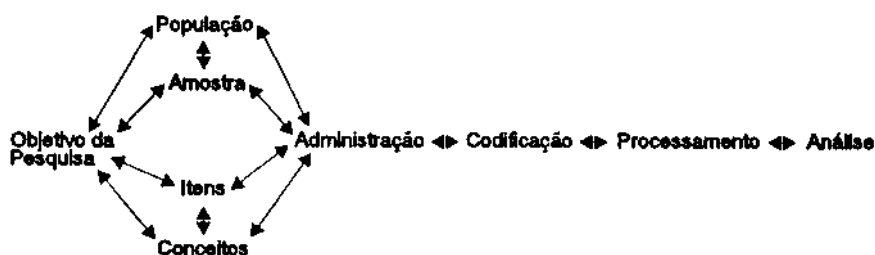
Fig. 15-1 — Estágios principais do *survey,* conforme Schuman e Kalton (1985, p.641).

Os componentes deste processo são intimamente interligados, sendo que existe uma influência mútua entre os mesmos. Inicialmente, comentaremos sobre este inter-relacionamento para, depois, focalizar na relação entre **objetivo da pesquisa, conceito** e **itens.**

Para fins de ilustração, usaremos o caso hipotético de um levantamento de dados utilizando a avaliação da qualidade de vida em um bairro por seus moradores. Como em qualquer levantamento de dados, existem

inicialmente duas vertentes a serem consideradas: o **objetivo** da pesquisa e a **população-alvo.** Um estudo desta natureza pode ter **objetivos** muito diferentes — por exemplo, a) uma avaliação dos serviços prestados por órgãos governamentais (água, luz, coleta de lixo, policiamento, educação, recreação), b) um estudo do bem-estar psicológico entre os moradores e o senso de comunidade e integração no bairro ou c) um levantamento da satisfação para com as moradias (tamanho, conservação, aluguel, estética.)

No que diz respeito à **população-alvo,** esta pode ser constituída por todos' os moradores do bairro ou por determinados subgrupos, como crianças, adolescentes, donas de casa, proprietários ou locatários. A relação entre objetivo e população é mútua: tanto o objetivo de um estudo implica uma determinada população-alvo quanto o interesse numa determinada população provoca certos objetivos para pesquisas.

Num segundo momento, o **objetivo** do estudo determinará os conceitos-chave que entram no levantamento de dados, da mesma maneira como a **população-alvo** implica a seleção de uma amostra na qual este levantamento de dados se realizará. A amostra sofrerá variação dependendo da população, sua constituição e tamanho. Se a população-alvo for constituída por todos' os moradores do bairro, é necessário que os subgrupos relevantes para o estudo sejam adequadamente representados. No caso de a população ser constituída por um determinado grupo, e.g., aposentados, o número e a maneira de selecionar os participantes do estudo sofrerão modificação. A questão da relação entre a **população-alvo** e a **amostra** dos participantes é assunto de livros especializados sobre amostragem. Este tema não será tratado neste capítulo, uma vez que uma introdução completa é fornecida por Kish (1965). A maioria dos textos de metodologia da pesquisa social incluem capítulos sobre amostragem (e.g., Ackoff, 1975; Aday, 1989; Babbie, 1992; Frankel, 1983; Judd, Smith, Kidder, 1992; Sudman, 1983).

Quanto à relação entre o **objetivo** de um estudo e os **conceitos** a serem pesquisados, ilustraremos com o caso hipotético da avaliação de um bairro por seus moradores. A avaliação de serviços implica conceitos como grau de satisfação, enquanto que um estudo do bem-estar psicológico entre os moradores poderia levar a conceitos tais como senso de comunidade, bem-estar consigo mesmo e com os vizinhos ou clima social entre moradores. Dependendo dos **conceitos** subjacentes, o conteúdo das perguntas ou **itens** varia. A forma de apresentação do conteúdo dos itens, por sua vez, varia em função do conteúdo, de características da população-alvo (e.g., nível educacional), bem como do modo de coleta de dados (e.g., entrevista pessoal, questionário auto-aplicado, etc.) Mais uma vez há que se enfatizar a relação mútua entre objetivo e conceito: tanto o objetivo da pesquisa determina os conceitos

a serem pesquisados nas diversas formas de itens quanto a possibilidade de desenvolver determinados itens para certas amostras influencia que objetivos poderiam ser atingidos.

Desta maneira há que se sublinhar que a seleção da população e amostra de um estudo não independe dos objetivos e conceitos (itens). Esta interdependência se reflete mais diretamente na maneira como se pretende aplicar o instrumento de levantamento de dados: 1) o grau de complexidade dos conceitos vai determinar tanto o número de itens quanto a forma dos mesmos; 2) as características da população-alvo determinam a complexidade dos conceitos que podem ser investigados, bem como a maneira da sua transformação em itens (perguntas) e sua administração; 3) o tamanho da amostra vai influenciar a maneira de administrar o instrumento. Se a composição da população-alvo ainda depende principalmente do objetivo da pesquisa, o tamanho da amostra é determinado tanto pelos conceitos a serem explorados quanto pelos recursos (tempo, dinheiro e recursos humanos) disponíveis. Esta disponibilidade de recursos influencia o planejamento da administração do instrumento, bem como a codificação das respostas, seu processamento e, eventualmente, as possíveis análises. Considerem-se os dois exemplos seguintes.

Uma pesquisa que visa determinar a satisfação com serviços básicos, como água, luz, limpeza, pode ter como população-alvo respondentes com telefone. Caso o pesquisador ainda tenha à sua disposição recursos para contratar entrevistadores, treiná-los e instalar dez linhas telefônicas e microcomputadores, seria possível levantar dados junto a um grande número de pessoas dentro de pouco tempo.

Já uma exploração de conceitos, como clima social, confiança mútua e solidariedade entre moradores de um bairro periférico, tendo como população-alvo pessoas analfabetas, exigiria outro tipo de instrumento, outra maneira de aplicação e administração, que resultaria num método de codificação, processamento e análise diferentes.

Em suma, embora este capítulo trate da construção do instrumento de levantamento de dados, o leitor deve se lembrar que sua realização dependerá da população-alvo, do tamanho da amostra, dos conceitos a serem explorados, bem como dos recursos disponíveis para a aplicação e processamento do instrumento.

CONSIDERAÇÕES GERAIS SOBRE UM INSTRUMENTO

Além de, como fizemos acima, enfatizar a interdependência entre população e amostra de um lado e objetivo e conceitos (itens) do outro, qualquer instrumento para levantamento de dados deve obedecer a uma

estrutura básica: a) identificação do pesquisador e legitimação dos objetivos da pesquisa e b) uma estrutura lógica das temáticas apresentadas na coleta de dados.

Identificação do pesquisador e legitimação dos objetivos da pesquisa

Considerando que o respondente de um levantamento de dados, quer via entrevista ou questionário, gasta seu tempo e faz algum esforço mental, uma primeira reflexão a ser feita é: **quem** quer algo **de quem,** numa determinada pesquisa? A disposição por parte do respondente em gastar tempo e fazer esforço, permitindo assim ao pesquisador obter os dados desejados, varia conforme a situação. Mencionamos aqui apenas alguns exemplos: confessionário (padre—fiel), interrogatório (policial—suspeito), declaração de renda (receita—contribuinte), vestibular e concurso (comissão de admissão—candidato), prova (professor—aluno), aconselhamento (psicólogo—cliente voluntário), manicômio (psiquiatra—cliente enviado), procura de emprego (funcionário de RH—solicitante), pesquisa de *marketing* (firma de pesquisa, podendo ou não oferecer brindes—respondente), pesquisa social universitária (pesquisador—'sujeito'). Neste trabalho, interessam os dois últimos exemplos, caracterizados pelo fato de o pesquisador não ter poder sobre o respondente e, conseqüentemente, precisar convencê-lo de que vale a pena responder.

O primeiro passo neste convencimento é uma boa apresentação, seja isto do instrumento ou da pessoa que o administra. Caso o questionário seja remetido via correio, vai acompanhado de uma carta de apresentação na qual consta a informação sobre quem está 'por trás' da pesquisa e para que serve a mesma. Em se tratando de uma entrevista, tanto pode ser que o entrevistador explique quem ele é, para quem trabalha, identificando-se de alguma forma (e.g., crachá) e, ainda, entregue alguma carta ao candidato a respondente. Como os primeiros momentos decidem sobre a disposição do respondente em continuar a ser 'importunado', é aí que qualidade e quantidade de informação sobre a pesquisa precisam se concentrar.

No caso da aplicação pessoal de um instrumento, o entrevistador, apropriadamente treinado, tem a oportunidade de explicar a) quem é responsável pela pesquisa, b) quais os objetivos e c) o que é que o respondente deve fazer. Mesmo assim, é vantajoso manter um contato prévio com os membros da amostra, avisando a chegada dos entrevistadores e explicitando os objetivos da pesquisa para facilitar sua recepção (Solórzano, 1991) ou aumentar a disposição em responder (Gouveia, 1991). Este 'aviso prévio' pode ser realizado através de uma carta ou até via campanha de *outdoors* ou anúncios na televisão. De qualquer modo, o entrevistador deve ter alguma maneira de se legitimar e ter algum texto explicativo preparado quando solicitar a colaboração do respondente em potencial.

Bom dia. Meu nome é José da Silva, trabalho para a firma *Pesquisas Sem Fim,* especializada em *enquêtes* de opinião pública. Estamos realizando uma pesquisa entre os moradores do bairro Nova Esperança. Solicito, assim, sua colaboração em responder algumas perguntas sobre a qualidade de vida no seu bairro.

Enquanto o entrevistador pode tirar qualquer dúvida neste contato inicial, deste modo convencendo respondentes vacilantes, num questionário remetido via correio, a carta precisa ser suficientemente atrativa para capturar a atenção inicial, para assegurar que a pessoa que a receba a leia com algum cuidado e adquira a informação necessária que a leve a prestar atenção ao questionário propriamente dito.

Esta identificação do pesquisador e legitimação dos objetivos da pesquisa é apenas o primeiro passo que permite que se comece o levantamento de dados. A necessidade de convencer o respondente a responder continua até o fim da entrevista ou do questionário, i.é, evitar que se desista no meio do processo depende tanto da forma quanto do conteúdo do instrumento.

Estrutura do Instrumento

A estrutura do instrumento assegura que todos os temas de interesse do pesquisador sejam tratados e que se mantenha o interesse do respondente em continuar. Uma primeira regra é que itens tratando de uma mesma temática devem ficar juntos, recebendo uma introdução que ajuda o respondente focalizar na temática a ser tratada:

Inicialmente, gostaria de saber da sua opinião sobre o sistema de segurança pública ...

Uma segunda consideração é que os conjuntos de itens que tratam de uma mesma temática devem seguir uma ordem. Esta ordem é do mais geral para o mais específico; do menos delicado, menos pessoal, para o mais delicado, mais pessoal. Desta forma, o último conjunto de itens trata das características socioeconómicas do respondente. Dependendo do assunto, a mesma regra 'do geral para o específico' pode ser aplicada à seqüência dos itens dentro de uma determinada temática. Ressalve-se entretanto que, em se tratando de um conjunto de itens que constituem uma escala, os itens devem ser misturados, para evitar que dois itens tratando essencialmente de um mesmo aspecto sejam apresentados um após o outro.

Na pesquisa hipotética entre moradores de um bairro realizada em forma de entrevista pessoal, a parte formal da coleta de dados poderia ser precedida por perguntas gerais sobre a situação do respondente na cidade e bairro:

Há quanto tempo mora nesta cidade? [Caso apropriado] Antes, morava onde? Em geral, está satisfeito em morar aqui?

Estas perguntas iniciais serviriam menos para obter informação do respondente e mais para estabelecer um relacionamento de confiança entre respondente e pesquisador. Entretanto, deve-se atentar para o fato de que estas perguntas terão que ser repetidas de maneira formal mais adiante, dentro da entrevista, enquadrando-se os dois primeiros itens dentro do conjunto da caracterização socioeconòmica e o terceiro na série sobre satisfação com o bairro.

Um terceiro princípio de organização do instrumento é que, na medida apropriada, este deve seguir uma ordem lógica. Usando a hipotética pesquisa sobre moradia como exemplo, pergunta-se inicialmente a cidade, depois sobre o bairro, a rua e o prédio onde o respondente mora. Além de progredir do geral para o mais específico, também vem se aproximando do respondente, i.é, uma pergunta sobre o relacionamento entre os moradores da cidade é menos pessoal, menos ameaçadora do que sobre o relacionamento do respondente com seu vizinho. Fazendo perguntas mais pessoais após ter estabelecido um certo nível de confiança entre entrevistador e entrevistado ajuda a eliciar respostas confiáveis (veja Günther, Brito, Silva, 1989).

Assim, perguntas pessoais sobre o respondente constituiriam o último conjunto:

Concluindo, gostaríamos de fazer algumas perguntas para melhor caracterizar os respondentes desta pesquisa...

Um erro comum de instrumentos é o de começar com perguntas pessoais, muitas vezes numa seção chamada 'identificação'. Em se tratando de pesquisa, não convém identificar o respondente. Antes, pelo contrário, geralmente há que se assegurar que a pesquisa não visa identificar indivíduos, mas que perguntas sócio-demográficas, como sexo, idade, educação, estado civil, composição da família, renda, tempo de moradia, etc, servem apenas para caracterizar a amostra. Perguntar o nome no início de uma entrevista pessoal pode facilitar trato interpessoal, mas, mesmo sem registrá-lo, pode contradizer qualquer afirmação de confidencialidade da entrevista.

Forma das perguntas

Uma primeira consideração diz respeito a perguntas abertas *versus* fechadas. A discussão é bastante extensa (veja Günther, Lopes, 1990). Pode-se, entretanto, sumariar a discussão nos seguintes termos: para uma pesquisa inicial, exploratória, não conhecendo a abrangência dos assuntos e trabalhando com relativamente poucos respondentes, perguntas abertas são mais apropriadas. Uma vez que se conhecem os tópicos geralmente

mencionados pelos respondentes acerca de uma dada temática, e especialmente quando se têm muitos respondentes e/ou pouco tempo, não existe razão para não usar perguntas fechadas. O argumento de que perguntas abertas dão mais liberdade ao responderte de se expressar é uma falácia: nas palavras de Sommer e Sommer (1991, p. 131), o uso de perguntas fechadas "mostra freqüentemente mais respeito à opinião das pessoas, deixando-as classificar suas respostas como positivas, negativas ou neutras, em vez do pesquisador fazer isto para eles".

Entretanto, da mesma maneira que perguntas abertas servem no início da entrevista para estabelecer um certo clima entre pesquisador e respondente, servem, no fim do levantamento, para capturar justamente aquelas opiniões não cobertas pelos itens fechados. Além de um 'apanhado final' ao concluir o questionário ou a entrevista, tais perguntas abertas podem ser feitas no fim de um conjunto de perguntas. Não somente servem para capturar qualquer pensamento do respondente não coberto pelas perguntas fechadas, mas servem para reforçar a percepção (essencial para o bom andamento da entrevista ou questionário) do respondente de que o pesquisador tem interesse na opinião pessoal e individual dele. Entretanto, há que se lembrar, perguntas abertas, especialmente em questionários auto-aplicados, exigem mais esforço por parte do respondente e, aumentando o custo da resposta, diminuem a probabilidade de completar e devolver o questionário.

DESENVOLVIMENTO DE ITENS

Partindo dos objetivos, formulam-se perguntas a serem respondidas através da pesquisa. As perguntas são transformadas operacionalmente em variáveis e indicadores. Para obter-se a informação, com a qual se trabalhará eventualmente na análise dos dados coletados no levantamento, criam-se itens a serem apreciados pelo respondente. É através da formulação de quesitos nos itens que se estabelece a relação entre o objetivo de uma pesquisa e os conceitos pesquisados. Através das respostas, tenta-se representar o grau de conceitualização do respondente acerca de um determinado assunto. Desta maneira, estamos nos aproximando de uma definição de medição: estabelecer uma correspondência entre eventos e símbolos, comumente numerais, de tal maneira que a variação entre os símbolos corresponde, geralmente de maneira linear, à variação entre os eventos. No caso da pesquisa social, 'evento' quer dizer "idéias, sentimentos, planos, crenças, bem como origem social, educacional e financeira", aquilo que, na definição de Fink e Kosecoff, está sendo coletado num levantamento de dados, enquanto que 'símbolo' é a apresentação de alternativas nos itens do instrumento para chegar à informação desejada. Para que tal correspondência ou medição seja fiel,

há que se atentar para três aspectos: erro, singularidade e representação. O evento é corretamente identificado e discriminado de outros eventos próximos? Cada evento é representado por apenas um símbolo, e cada símbolo representa apenas um evento? Qual a maneira com que os símbolos (numerais) representam os eventos?

Nas ciências sociais, diferenciam-se quatro níveis de correspondência entre eventos e símbolos, i.é, entre quatro tipos de escalas, sumariados na Tabela 15-1. Far-se-á referência às categorias desta tabela nas considerações a seguir. Pasquali (capítulo 1, neste livro) apresenta com mais detalhe esta correspondência entre eventos e símbolos.

Tabela 15-1 — Características de escalas nas ciências sociais

TIPOS DE ESCALA	CARACTERÍSTICAS DA ESCALA	EXEMPLOS	CARACTERÍSTICAS FORMAIS
Nominai	Números e símbolos são utilizados somente para identificar pessoas, objetos ou categoria	Placas de carro, cor de cabelos. local de nascimento, estado civil	Equivalência ■ - '
Ordinal	Características podem ser ordenadas numa dimensão subjacente	Ordem de chegada, ordem de preferência, *status* social, escala Likert	Além da anterior, um ¡tem maior do que o outro ' > '
Intervalar	Características não somente podem ser ordenadas numa dimensão subjacente, mas intervalos têm tamanho conhecido e podem ser comparados	Escalas de Thurstone, escala de Likert, estimativas de distâncias, temperaturas em oC	Além das anteriores, operações aritméticas nas diferenças entre os números representando eventos
Razão	Alem das características da escala anterior, ainda existe um ponto zero absoluto	Salário.tamanho. tempo gasto com uma tarefa	Além das anteriores, operações aritméticas nos próprios números

Fonte: Siegel (1956) e Sommer e Sommer (1991)

Considerações gerais ao desenvolver itens

Linguagem

Quanto à linguagem usada na formulação dos itens, atenta-se, inicialmente, para a compreensão dos mesmos pela população-alvo da pesquisa. Abreviações, gíria profissional, termos regionais devem ser evitados, da mesma maneira que termos especiais ou sofisticados para uma determinada população-alvo.

Ambigüidade

Relacionado à questão da linguagem em si está o problema da ambigüidade. O respondente está entendendo aquilo que o pesquisador está perguntando?

Viés e ênfase

Finalmente, a escolha das palavras pode dirigir as respostas. Quando se pergunta sobre utilização de áreas comuns num bloco de residência, pode-se indagar se algo deve ser 'proibido', 'não permitido', 'evitado' ou 'impedido'. Da mesma maneira que os avisos 'proibido estacionar' e 'pede-se não estacionar* provocam comportamentos diferentes, o número de respondentes que concordam com um item que contém a palavra 'proibir' e 'não permitir' varia (Schuman, Presser, 1981).

Assim, convém sempre realizar um estudo piloto para verificar **se** e **como** as perguntas estão sendo entendidas pelos respondentes.

Apresentação dos itens

A apresentação dos itens de um levantamento de dados pode ser conceitualizada como a apresentação de um estímulo do qual se espera alguma resposta, algum comportamento que, por sua vez, precisa ser registrado de alguma forma. Desta maneira, há potencialmente três atores envolvidos direta ou indiretamente: quem administra o instrumento, quem responde ao instrumento e quem transcreve a informação registrada no instrumento para o processamento e análise dos dados.

No caso de um questionário auto-aplicado, tal estímulo é visual. O enunciado do item geralmente é em forma de algum texto, através do qual se formulam as perguntas e/ou instruções. Já as alternativas podem ser frases ou palavras, ao lado das quais se responde marcando um ✓, uma série de palavras ou números ou símbolos [e.g., ☺ ou ☹] que se risca com
um X.

No caso de uma entrevista via telefone, o estímulo é auditivo, enquanto que numa entrevista pessoal, frente a frente com o participante, os itens, os estímulos, podem ser apresentados de todas as maneiras, sendo auditivos, visuais, tácteis ou olfativos, mostrando-se fotografias, objetos, essências, entre os quais o respondente aponta um ou outro, conforme a instrução.

Uma segunda consideração na apresentação dos itens é a de evitar estimulação concorrente. No caso do instrumento auto-aplicado, é impossível controlar o ambiente onde o respondente preenche o questionário. Já numa interação pessoal, pode-se controlar tal estimulação concorrente através da escolha do local. Não se deve esquecer que o próprio compor-

tamento do entrevistador pode representar uma estimulação concorrente: imagine-se este manuseando uma prancha com o instrumento, lápis, três fotos entre as quais o respondente deve escolher uma, mais o material já usado bem como aquele a ser usado posteriormente. Se o entrevistador não for bem treinado, corre-se o risco de confundir o respondente antes deste fornecer alguma informação válida. Escolhendo para a aplicação um local calmo, de acesso restrito, com uma boa mesa, reduzem-se interferências indesejadas.

Uma terceira consideração diz respeito aos até três atores envolvidos na administração de um instrumento de levantamento de dados. O primeiro, que apresenta os itens, precisa ser bem treinado para assegurar que a estimulação seja a mais semelhante em todos os contatos com os respondentes. A opinião emitida pelo respondente deve representar sua reação às alternativas apresentadas, não a quem as apresentou. Dentro de certos limites, isto pode até ser automatizado, quando os itens são apresentados via computador ou gravados, no caso de entrevistas via telefone. Obviamente, quanto mais estandardizada a apresentação dos estímulos, i.é, dos itens, mais se perde o elemento 'humano' de uma interação, aspecto justamente que leva em conta a situação e o estado de espírito da situação. A preocupação com uma maior estandardização da apresentação dos itens acontece em levantamentos de dados que a) se assemelham a testes, b) solicitam informações mais objetivas ou c) coletam dados entre muitos respondentes que precisam ser apurados de maneira rápida.

Considerando o segundo ator, o respondente, a maneira de apresentar os estímulos, os itens, deve corresponder às suas habilidades, sejam estas intelectuais (saber 1er) quanto físicas (ver, ouvir, discriminar cheiro ou gosto). O que foi dito a respeito da compreensão da linguagem acima estende-se ao uso de símbolos e fotografias. O ☻ é entendido e interpretado como 'concordância'? Aquela foto, caso escolhida pelo respondente como representando um escritório mais confortável, permite a inferência de que o respondente é dinâmico?

Quanto ao modo de registrar as respostas de um levantamento de dados, convém pensar, desde o planejamento da pesquisa, no processamento e na análise dos dados. Enquanto respostas a perguntas abertas precisam ser decifradas, transcritas, codificadas, digitadas e verificadas quanto à sua consistência com as demais respostas (a proverbial mulher de 12 anos que relatou dois abortos e três gravidezes), o uso de um computador na apresentação dos itens bem como no registro das respostas facilita a apuração e assegura maior fidedignidade dos dados. Questionários que contêm apenas perguntas objetivas podem ser acompanhados de um cartão especial para o registro das respostas que, por sua vez, pode ser lido mecanicamente. No caso da transcrição, por alguém, dos dados registrados numa folha de resposta ou no próprio

questionário, deve-se pensar nas capacidades de quem transcreve ou digita. Antes de o instrumento ser entregue ao digitador, deve ter sido 'limpo' de tal maneira que não requeira mais julgamento por parte dele (e.g., o respondente marcou um três ou um quatro naquele item?). O *layout* do questionário deve favorecer orientação no que diz respeito ao seqüenciamento da informação a ser transcrita. Se há texto como resposta a perguntas abertas, não somente deve ser legível, mas deve, também, ser claro o que deve ser transcrito — O texto todo? Apenas uma parte? Que parte?. O *layout* e as instruções ao respondente devem facilitar a leitura das respostas pelo digitador; por exemplo, no caso de itens de escolha múltipla, deve-se apresentar números em vez de palavras ou letras e, ainda, pedir que o respondente os **circule** em vez de **marcar com X.**

Classificação de itens conforme tipo de escala

Voltamos à situação de um questionário padrão, para considerar diferentes tipos de itens. Apresentamos os mesmos a seguir, em termos de tipo de escala subjacente às alternativas de respostas. Consideramos, neste momento, apenas perguntas fechadas.

Escala nominal

Conforme os exemplos dados na Tabela 15-1, numa escala nominal utilizam-se números ou símbolos somente para identificar pessoas, objetos ou categorias. Exemplos para as ciências sociais seriam local de nascimento, sexo, estado civil, bem como certos atributos (como cor de cabelo) ou uso de aparelhos (como óculos ou bengala). A forma de apresentar estes itens é a seguinte:

Qual o estado civil de V.Sa.?

Solteiro(a) 1
Casado(a) 2
Vivendo maritalmente 3
Desquitado(a) 4
Divorciado(a) 5
Separado(a) 6
Viúvo(a) 7
Outro 8

Apontamos para alguns aspectos deste item. Primeiro, mesmo ao se preparar um instrumento para auto-aplicação, deve-se pensar em um diálogo com o respondente. Contrariamente a uma declaração de renda ou ficha de procura de emprego, convém estabelecer um bom *raport* com o respondente. Assim, uma frase como 'qual o estado civil de Va.Sa.?' (ou, conforme relacionamento com a população-alvo, 'o seu') soa melhor do que solicitar simplesmente 'estado civil'. Conforme o que nós sabemos sobre a população-alvo, mais ou menos alternativas podem ser apresen-

tadas. Muitas vêzes, 'solteiro, casado, outro[1] é suficiente. O importante é que as opções sejam a) mutuamente exclusivas e b) cubram todas as alternativas. Uma outro maneira de formular alternativas do estado civil é:

Nunca casado(a)	1
Sempre casado(a), i.é, casado(a) e nunca divorciado(a)	2
Divorciado(a)...	3
Recasado(a)	4

Dependendo do objetivo da pesquisa para o qual um determinado item está sendo utilizado, o primeiro ou o segundo exemplo do item 'estado civil' pode ser mais apropriado. Entretanto, da mesma maneira como o leitor pode ter reagido inicialmente ao segundo exemplo, sem dúvida, a maioria dos respondentes reagiria ainda mais. É um exemplo concreto de distinguir entre o conceito subjacente que se quer analisar numa determinada pesquisa (i.é, as quatro categorias do segundo exemplo) e o que se pode, do ponto de vista prática e conceitualmente factível, perguntar à maioria dos respondentes.

Escala ordinal

Conforme os exemplos dados na Tabela 15-1, numa escala ordinal os números ou símbolos, além de identificarem pessoas, objetos ou categorias, os ordenam numa dimensão subjacente. Exemplos para as ciências sociais seriam hierarquização de preferência entre objetos, *status* social ou itens de uma escala Likert. A forma de apresentar estes itens é a seguinte:

Como você sabe, a Prefeitura está lançando um programa de recuperação dos bairros da cidade. Solicitamos que indique entre as medidas listadas a seguir qual a que deve ser realizada primeiro, qual a segunda a ser realizada, qual a terceira e qual a quarta:

	N° de ordem de realização
Consertar a infra-estrutura (calçadas, iluminação pública, etc.)	_____
Consertar as vias e sinalização de transito...........................	_____
Construir/consertar escolas do bairro	_____
Construir/consertar os postos de saúde do bairro	_____
Outros, quais?	_____

A tarefa do respondente é a de marcar a ordem de importância no espaço indicado. Para cada um dos quatro itens ou, posteriormente, variáveis, pode-se determinar uma distribuição de freqüência: Quantas vezes 'sinalização de trânsito' foi mencionada como sendo a mais importante, a segunda, a terceira e a quarta medida. Ainda, é possível sumariar estes dados indicando quantas vezes cada um dos itens (no caso, medidas da prefeitura) foi mencionado como o mais importante.

Desta maneira, os valores modais e medianos podem ser calculados, mas não a média. Um outro exemplo de escalas ordinais é apresentado pela escala de valores de Rokeach (1973; Günther, 1981). A tarefa do responderte consiste em ordenar 18 valores segundo a sua importância para si. Apesar de Rokeach ter sugerido o uso de etiquetas para forçar o responderte a atribuir um determinado valor hierárquico a cada valor ou item no lugar de espaços como no exemplo anterior, os dados apresentam uma série de dificuldades para análise. Moore (1975) mostrou que, apresentando os 18 valores da escala de Rokeach na versão original ou como 18 itens de uma escala tipo Likert (veja a seguir), obtêm-se, essencialmente, os mesmos resultados, sendo que a escala Likert é mais fácil de responder e analisar.

Escala Likert

Por ter sido citada como exemplo de escala ordinai e escala intervalar, consideremos a escala tipo Likert separadamente. Ela é a mais utilizada nas ciências sociais, especialmente em levantamentos de atitudes, opiniões e de avaliações. Nela apresenta-se um determinado número de alternativas para julgar um enunciado, conforme o seguinte exemplo:

> Inicialmente, gostaríamos de saber o que os moradores do bairro acham sobre os serviços oferecidos pela Prefeitura. Para cada serviço a seguir, indique, por favor, se você o avalia como (1) muito ruim, (2) ruim, (3) razoável, (4) bom ou (5) muito bom. Para isto, faça um **círculo** em volta do número que melhor representa sua avaliação:
>
> Segurança pública (pode-se andar na rua à noite sem medo)
>
> Muito ruim 1
> Ruim.. 2
> Razoável 3
> Bom 4
> Muito bom 5
>
> [seguem mais itens]

Conforme foi mencionado anteriormente, uma série de itens tratando de um mesmo assunto recebe uma introdução em conjunto que orienta o respondente: apresenta a temática do conjunto, orienta quanto às alternativas e dá instruções concretas, e.g., faça um **círculo** em volta do número que melhor representa sua avaliação'. Neste exemplo, foi apre-

* Os 18 valores são impressos em etiquetas. O respondente cola as etiquetas em determinados espaços, conforme a importância atribuída. Evita-se, desta maneira, a possibilidade de empates.

sentado um número **ímpar** de alternativas. Embora geralmente se use quatro ou cinco alternativas nas escalas tipo Likert, encontram-se também itens com duas, três ou até nove alternativas. A decisão principal a ser tomada pelo pesquisador é se deve usar um número par ou ímpar de alternativas, i.é, deixar para o respondente a opção de não se comprometer, podendo marcar um ponto no meio de uma escala com número ímpar de alternativas. Hã que se diferenciar entre não saber opinar sobre um determinado tema e não querer se comprometer. Quando existe a possibilidade de o respondente não ter condições de responder, deve-se deixar uma alternativa explícita de 'não sei'. Tal alternativa, entretanto, não deve ser tratada, posteriormente, como ponto neutro no meio da escala, uma vez que 'indefinido - não saber' é diferente de uma atitude 'indefinido = - mais ou menos' no meio de uma escala! Quanto à segunda possibilidade, de o respondente não querer se comprometer, provavelmente deixaria o item em branco, não sendo conveniente estimular tal comportamento apresentando uma alternativa como 'mais ou menos'.

Dependendo do tema subjacente, as alternativas podem, além da dimensão bom—ruim, seguir a dimensão concordo—discordo. Muitas vezes, é apenas uma questão de reformular o estímulo/item para mudar as categorias de resposta de bom—ruim para concordo—discordo. Obviamente, convém formular os itens de um conjunto de tal maneira que todos eles possam ser respondidos na mesma dimensão (veja Sommer, 1991). Caso a dimensão concordo—discordo seja utilizada, também é boa prática que alguns dos itens sejam invertidos, para que ora 'concordo' ora 'discordo' implique uma atitude favorável com o conceito subjacente ao conjunto de itens, conforme demonstrado nestes dois itens da Escala de Ambiente de Trabalho (EAT) de Moos (1987):

> A seguir você encontrará uma série de afirmações a respeito do seu ambiente de **trabalho.** Solicito que indique quais destas afirmações se aplicam a você e quais delas não se aplicam. Caso a questão se aplique a você, faça um circulo em volta da palavra SIM. Caso a questão não se aplique, faça um circulo em volta da palavra NÃO.
>
> O trabalho realmente apresenta desafios..................... Sim Não
> Muitas pessoas parecem apenas passar o tempo........ Sim Não

A concordância ['sim'] com o primeiro item e a discordância com o segundo implica uma atitude favorável na dimensão envolvimento com o trabalho.

Tratamos a escala Likert separadamente, uma vez que é mencionada tanto como exemplo da escala ordinal quanto intervalar. Estritamente, as alternativas de um item da escala tipo Likert representam uma medição no nível ordinal. Os valores numéricos (e.g., 1, 2, 3 e 4) associados às

alternativas 'discordo fortemente', 'discordo', 'concordo', e 'concordo fortemente' não permitem operações formais além de ' > '. Entretanto, somando as respostas a vários itens e/ou tendo respostas de muitos respondentes, pode-se supor que as diferenças entre as alternativas se ajustam de tal maneira que permitem operações aritméticas reservadas a escalas intervalares.

Escala intervalar

Conforme consta na Tabela 15-1, numa escala intervalar as características não somente podem ser ordenadas numa dimensão subjacente, mas os intervalos têm um tamanho conhecido e podem ser comparados, enquanto que itens de uma escala tipo Likert com alternativas na dimensão concordo—discordo mostram uma certa dificuldade em ser considerados como uma escala de intervalo. Isto pode ser contornado apresentando-se uma verdadeira escala ou régua:

> Utilizando a seguinte escala, que varia de muito ruim a muito bom, avalie os serviços prestados pela Prefeitura, riscando o ponto na mesma que melhor representa sua avaliação:
>
> Serviço de Limpeza Urbana   Muito I______________________ I Muito
> Ruim   Bom

Enquanto este tipo de representação visual de um escala pode 'medir' com maior certeza o grau de avaliação, é muito difícil de apurar, uma vez que se precisa de uma régua milimetrada para determinar exatamente onde o respondente marcou sua resposta nesta escala. Um outro exemplo clássico de uma escala intervalar é a utilizada por Milgram (1974) para determinar o grau de obediência dos participantes às instruções nas suas pesquisas. Ostensivamente, o participante aplicava choques elétricos que variavam entre 15 e 450 volts. O grau de obediência correspondia à voltagem onde o participante se recusava a continuar aplicando mais choques, i.é, quanto mais baixa a voltagem, menos obediente.

Além das escalas tipo Likert, são as do tipo Thurstone que satisfazem as exigências da escala intervalar. Na base de estudos pilotos, estabelece-se o valor atitudinal de cada item que será atribuído ao respondente, caso concorde com a afirmação. A apresentação dos itens não difere da do exemplo anterior da Escala de Ambiente de Trabalho.

Escala de razão

Exemplos da escala de razão utilizada nas ciências sociais são salário ou tempo gasto com uma determinada tarefa. A apresentação dos itens reverte a perguntas abertas:

Considerando o seu tempo livre e de recreação, solicitamos que indique:

V.Sa. é membro de algum clube social? .... Sim Não
Caso sim,
— quanto tempo gasta, em média, por semana, neste clube? __ horas
— além da mensalidade do clube, quanto dinheiro gasta, em média, por mês, em atividades relacionados a este clube? .... R$ _______

Neste exemplo, a primeira resposta (sim ou não) representa urna medição em nível nominal, enquanto que as duas a seguir representam medições numa escala de razão.

Considerações gerais sobre os níveis de medição

Do ponto de vista da análise estatística, medições em nível nominal podem ser convertidas em escalas intervalares. Quando existem apenas duas alternativas codificadas, como 0 e 1, não existe necessidade para operações adicionais. Exemplos para isto são perguntas solicitando respostas como sim *versus* não, presente *versus* ausente, ou sexo. Já itens oferecendo mais do que duas alternativas, por exemplo, estado civil, região de nascimento ou afiliação religiosa, podem ser convertidos em uma série de alternativas binarias através do processo de codificação *dummy,* permitindo então operações estatísticas reservadas a escalas intervalares **e** de razão (veja, por exemplo, Pedhazur, 1982, Tabachnick e Fidell, 1983, ou Stevens, 1986).

REFERÊNCIAS BIBLIOGRÁFICAS

ACKOFF, R.L *Planejamento de pesquisa social.* São Paulo: E.P.U., 1975.

ADAY, L.A. *Designing and conducting health surveys.* San Francisco (CA): Jossey-Bass, 1989.

BABBIE, E. *The practice of social research* 6.od. Belmont (CA): Wadsworth, 1992.

FINK, A, KOSECOFF, J. *How to conduct surveys:* a step-by-step guide. Beverly Hills (CA): Sage, 198S.

FRANKEL, *M.* Sampling theory. In: ROSSI, P.H., WRIGHT, J.D., ANDERSON, A.B. (Eds.). *Handbook of survey research.* Orlando (FL): Academic Press, 1983. p.21-67.

GOUVEIA, V.V. *Survey pelo correio sobre percepção de barulho ambiental:* variáveis de técnica e de conteúdo. Brasília, 1991. Dissertação (Mestrado) — Universidade de Brasília.

GÜNTHER, H. Uma tentativa de traduzir e adaptar a escala de valores Rokeach para uso no Brasil. *Arquivos Brasileiros de Psicologia,* Rio de Janeiro, v.33, n.3, p.58-72,1981.

GÜNTHER, H., BRITO, S.M.O., SILVA, M.S.M.M. Relação entrevistador—entrevistado: um exemplo de técnica de entrevista com grupos marginalizados: meninos na rua *Psicologia: Reflexão e Crítica,* Porto Alegre, v.4, n.1/2, p.12-23, 1989.

GÜNTHER, H., LOPES JR., J. Perguntas abertas vs. perguntas fechadas: uma comparação empírica. *Psicologia: Teoria e Pesquisa,* Brasília, v.6, p.203-213,1990.

JUDD, CM., SMITH, E.R., KIDDER, L.H. *Research methods in social relations.* 6.ed. Fort Worth (TX): Holt, Rinehart and Winston, 1992.

KISH, L. *Survey sampling.* New York: Wiley, 1965.

MILGRAM, S. *Obedience to authority,* an experimental view. New York: Harper and Row, 1974.

MOORE, M. Rating versus ranking in the Rokeach Value Survey: an Isreali comparison. *European Journal of Social Psychology,* Sussex, v.5, p.405-408, 1975.

MOOS, R.H. *The social climate scales:* a user's guide. Palo Alto (CA): Consulting Psychologists Press, 1987.

PEDHAZUR, E.J. *Multiple regression in behavioral research:* explanation and prediction. New York: Holt, Rinehart and Winston, 1982.

ROKEACH, M. 77ie *nature of human values.* New York: Free Press, 1973.

SCHUMAN, H., KALTON, G. Survey methods. In: LINDZEY, G., ARONSON, E. (Eds.). *Handbook of social psychology.* 3.ed. New York: Random House, 1985. v.1, p.635-697.

SCHUMAN, H., PRESSER, S. *Questions and answers in attitude surveys:* experiments on question form, wording, and context New York: Academic Press, 1981.

SIEGEL, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

SOLORZANO, I.M. *Padrões de resposta e taxa de participação em levantamentos de campa* aplicação ao problema do ruído urbano. Brasília, 1981. Dissertação (Mestrado) — Universidade de Brasília.

SOMMER, R. Literal vs. metaphorical interpretations of scale terms: a serendipitous natural experiment. *Educational and Psychological Measurement,* Durham (NC), v.51, p. 1009-1012, 1991.

SOMMER, B., SOMMER, R. *A practical guide to behavioral research:* tools and techniques. New York: Oxford University Press, 1991.

STEVENS, J. *Applied multivariate statistics for the social sciences.* Hillsdale (NJ): Lawrence Erlbaum, 1986.

SUDMAN, S. Applied sampling. In: ROSSI, P.H., WRIGHT, J.D., ANDERSON, A.B. (Eds.). *Handbook of survey research.* Orlando (FL): Academic Press, 1983. p.145-194.

TABACHNICK, B.G., FIDELL, L.S. *Using multivariate statistics.* New York: Harper and Row, 1983.

CAPÍTULO 16

EVALUACIÓN CONDUCTUAL: UNA EXPERIENCIA DE INTEGRACIÓN

Eleonora Vivas
Universidad Simón Bolivar
Caracas, Venezuela

En el presente capítulo se plantea la experiencia de integración y los elementos que han caracterizado a la Evaluación Conductual que se realiza ligada a la investigación y docencia, en el marco del Postgrado en Psicología de la Universidad Simón Bolivar (Venezuela). En primer lugar se hace referencia a los diferentes niveles de integración donde se considera que ésta ha ocurrido: a nivel conceptual, a nivel metodológico y a nivel práctico. Se describe cada uno y, finalmente, se plantean las perspectivas futuras de la evaluación en nuestro contexto, de acuerdo a una proyección de la actividad actual.

¿QUE ES LA EVALUACIÓN CONDUCTUAL?

La Evaluación Conductual, como un nuevo enfoque hacia la evaluación, se ha ido desarrollando como producto de la evolución natural de un modelo de evaluación e investigación surgido de un enfoque fundamentalmente de caracter conductual y de naturaleza bastante comprehensiva. Este se basa, en principio, en una conceptualización del funcionamiento del ser humano como individuo inmerso en un medio donde juega tanto el papel de actor como de receptor. Es un modelo que unifica el diseño completo para recabar una información tanto comprehensiva como compleja sobre los diferentes elementos objeto de la evaluación, como la manera de recolección de los datos y su análisis e interpretación.

Los principales elementos del modelo, son los siguientes: 1. el sujeto-meta o grupo objeto de estudio; 2. el ambiente donde se desenvuelve el sujeto; 3. las relaciones funcionales entre los elementos antes mencionados. Por otra parte, la organización temporal de la búsqueda de información se establece de antemano dictada por las exigencias del mismo modelo. Este plantea la necesidad de verificación constante de la información, estableciendo así un diseño que si bien puede o no ser de

medidas repetidas, sí contempla al menos una evaluación inicial, otra final y un seguimiento, todos realizados de manera comparable. Y la forma de cuantificación objetiva y precisa forma también parte de las característiscas de este modelo, el cual es aplicable a una gran diversidad de conductas y situaciones.

Me referiré en el presente capítulo a la experiencia que en materia de evaluación hemos tenido aquellos que trabajamos en el área cognitivo-conductual del Departamento de Ciencia y Tecnología del Comportamiento de la Universidad Simón Bolívar, en Venezuela. Para ello haré referencia **a** las bases conceptuales sobre las que descansa la evaluación en el presente, como producto de una evolución de la actividad investigativo-docente, a las características que ésta tiene y, sin duda, a cómo ésta fué desarrollando. Finalmente presentaré nuestras ideas acerca de cuales serán los derroteros que podría seguir la evaluación, de continuar el trabajo docente y de investigación, por la senda que actualmente transita.

Desarrollo

Haciendo un poco de historia, la experiencia en Evaluación Conductual comenzó en el área de la docencia como un apoyo y ampliación de los aspectos metodológicos de una asignatura básica de Análisis Experimental de la Conducta, tal y como se utilizaba en el laboratorio. En este sentido se inscribía dentro de una conceptualización de tipo conductista radical, basado fundamentalmente en el modelo Skinneriano. Si bien en principio se describía la metodología que se utilizaba en los experimentos de la investigación básica, no iba acompañada de una praxis que le proporcionara un referente directo y con experiencia de primera mano. En esta primera etapa se recurrió sólo a las referencias bibliográficas, pues no estaba dentro de los planes de desarrollo de nuestro grupo hacer instalaciones de laboratorio para este tipo de investigación.

Posteriormente se introdujeron los **registros observacionales** de tipo conductual, realizados en el ambiente natural, y los diseños de un solo sujeto, tipo reversión, criterio cambiante y L-B múltiple. Esto respondió al mayor énfasis que al Análisis Conductual Aplicado se dio en ese momento. Los aspectos metodológicos que suscribía el Análisis Conductual Aplicado para la epoca constituyeron una alternativa viable en la investigación aplicada, sobre todo como contraste ante una "metodología tradicional" avasallante y poderosa. De acuerdo a las etapas señaladas por Fernández Ballesteros (1987), nosotros "llegamos" a la Evaluación Conductual en su etapa de 'constitución'. Esta la ubica Fernández Ballesteros en la década entre 1965 y 1975, caracterizada por la presentación de una alternativa funcional, o analítico-conductual, a la

evaluación psicológica. Hubo una consonancia con Kanfer y Saslow (apud Fernández Ballesteros, 1991) quienes plantean que la evaluación debe dirigirse a "la descripción de la conducta problemática, los tactores que la controlan y el medio a través del cual puede ser modificada".

Con este énfasis aplicado, la evaluación estuvo más que todo circunscrita a la obervación directa de conductas, realizadas especialmente en el ambiente natural del sujeto o sujetos a estudiar. Igualmente, fueron de los diseños de investigación intra-sujeto los que predominaron en las experiencias realizadas en las actividades prácticas y las intervenciones, tipo investigación, que se implementaron tanto en escuelas como en otros contextos, con niños al igual que con adultos. Esta fue parte de la reacción de la corriente conductual a la investigación de tipo "tradicional" y basada exclusivamente en diseños de grupos.

Sin embargo, este modelo basado en el Análisis Conductual Aplicado se fué quedando corto, pues ya la interacción entre la docencia y la investigación estaba generando necesidades específicas en relación al tipo de información del sujeto que era necesario considerar para obtener una perspectiva más completa de su funcionamiento. También, y por qué no decirlo, siguiendo la tendencia general en el área, se introdujeron aspectos menos observables, pero necesarios para una evaluación y análisis menos parcelados del individuo. Los aspectos internos, no observables del sujeto, era necesario tomarlos en cuenta. Nos referimos, entre otros, a la introducción del **auto-informe** o auto-reporte como una fuente de información importante, que aporta datos no accesibles de otra manera. El trabajo en el área del auto-control planteó la inminente necesidad de contar con la experiencia interna del sujeto y a la vez considerar al auto-informe como una alternativa válida capaz de reflejar una parte importante de la conducta. Adicionalmente al trabajo con auto-control se incursionó en el área de la Psicología de la Salud, la cual necesariamente debe basarse en información por parte del sujeto tanto referida a su experiencia vital como a sus concepciones acerca de tópicos particulares. En este sentido se ha hecho uso tanto de escalas ya elaboradas y probadas por otros autores para la exploración de variables tales como apoyo social, estrés laboral, auto-control, etc.; como otras diseñadas *ad hoc* para estudiar aspectos específicos de interés.

Este tipo de instrumento de evaluación proporcionó datos recopilados de una manera sistemática, y aunque provenientes de instrumentos validados, no fueron utilizados con referencia a normas estandarizadas, como es el uso de los típicos tests psicométricos. La información obtenida a través de estas escalas cumple un papel de tipo más bien descriptivo.

Adicionalmente, la necesidad de tener acceso a datos de tipo **psicofisiològico** que brindasen un cuadro más completo del individuo, introdujo dentro de la evaluación otro componente importante que

completa el cuadro correspondiente a los tres canales de respuesta. En este sentido se pueden señalar las evaluaciones de ritmo cardíaco, de tensión arterial, de reactividad cardiovascular y otras que complementan una evaluación más integral del individuo, las cuales son de gran relevancia para el estudio de factores psicofisiológicos ligados a respuestas emocionales y enfermedades cardiovasculares. Este es sólo un ejemplo de la aplicabilidad y, más que todo, la necesidad de contar con evaluaciones que brinden información de este canal de respuesta que en algunos casos, dependiendo del objeto de estudio, no se puede dejar de lado.

Y como marco referencial para la acción del individuo, también el **ambiente** donde éste se desenvuelve es tomado en cuenta desde diferentes puntos de vista, tanto en lo relativo al aspecto del entorno físico y social como en cuanto a la percepción que tiene el individuo de él. En la Figura 16-1 se muestra el desarrollo del modelo de evaluación.
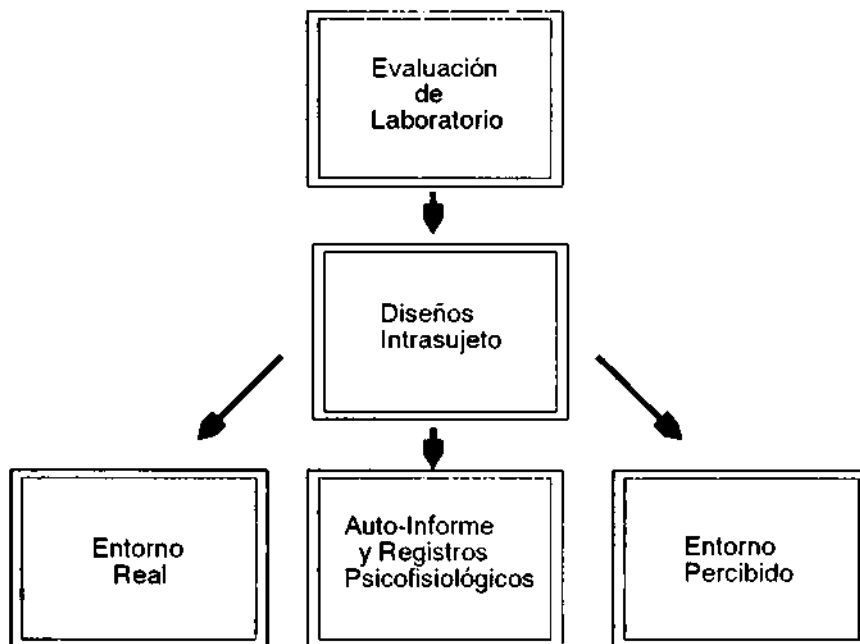
Fig. 16-1 — Desarrollo del modelo de evaluación.

En la actualidad la interacción que se mencionaba al principio, docencia-investigación, se ha invertido en cuanto a cuál de los dos elementos plantea las necesidades, y se ha generado más bien una interacción investigación-docencia. Esto es, el tipo de evaluación que se requiere en la investigación dicta más las pautas de lo que se va a abordar en la docencia.

Estas características de desarrollo ilustran una de las formas de **integración** que provienen de la praxis de dos actividades que a veces suelen desenvolverse de manera separada. Nuestra experiencia ha planteado una complementahdad que llevó a la integración.

Características actuales — niveles de integración

Integración conceptual y empírica

Toda esta metodología y procedimientos evaluativos van enmarcados dentro de una conceptualización que responde a las bases teóricas que la sustentan (ver Figura 16-2). De allí que, en principio se toma en cuenta, y de hecho se parte, de la posibilidad del ser humano de simbolizar, pensar y planificar por adelantado, se toma en cuenta la capacidad auto-reflexiva y auto-regulatoria del individuo (Bandura, 1986, p.18-21). Igualmente, y siguiendo esta conceptualización, es el determinismo recíproco de Bandura el cual se ha adecuado mejor a las expectativas que consideramos apropiadas para una evaluación integral del ser humano. Por supuesto, dentro de lo que cabe tomando en cuenta las restricciones que impone la realidad práctica.
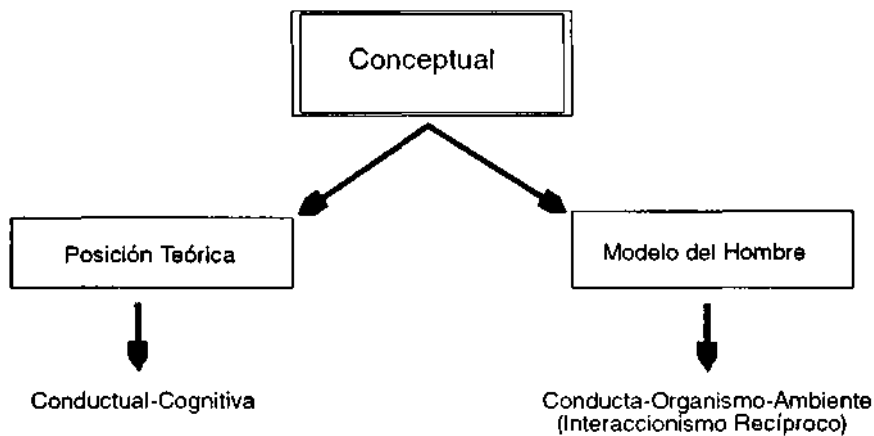
Fig. 16-2 — Nivel de integración conceptual

Este modelo del hombre se podría ejemplificar, en lo referente a la evaluación, en los términos que se muestra en la Figura 16-3, y basándonos en el esquema de la reciprocidad triàdica.

Este modelo esta constituido por elementos que interactuan reciprocamente: el componente "conductual", el cual se basa en la conducta abierta y manifiesta, observable por cualquier persona, es y ha sido abordado desde el punto de vista de la evaluación a través de la observación
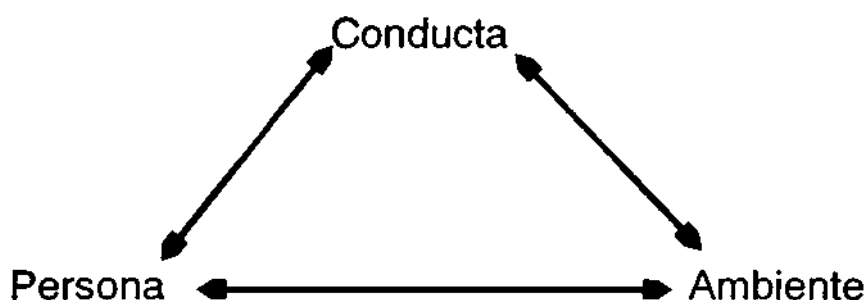
Fig. 16-3 — Bases conceptuales: esquema de reciprocidade triàdica de Bandura.

directa; el componente "persona", el cual representa el afecto y el pensamiento, y es evaluado a través del auto-reporte; y el "ambiente", el cual se refiere a aquel entorno donde se desenvuelve la acción del individuo, se evalúa tanto el ambiente real como el percibido.

Todos estos componentes se relacionan entre sí de manera probabilistica, interactiva, como determinantes unos de otros y, como señala Bandura, la reciprocidad se refiere a la acción mutua entre los factores. Sin embargo, esta acción no es ni simétrica ni de igual fuerza en la causación recíproca. De allí que en un momento dado tenga más preponderancia un componente sobre otro.

Tomando estos elementos en consideración, se ha implementado una forma evaluativa, que contemple de la manera más integral posible todos los aspectos mencionados.

Otro aspecto de importancia a destacar es la importancia que se le asigna a tomar en cuenta tanto los **tres componentes** de respuesta (conductual, cognitivo y fisiológico), como los **tres elementos de la tríada** (conducta, persona y ambiente). Todos estos elementos son evaluados a través de procedimientos y medidas complementarias. Acá se evidencia la adopción de un modelo integrativo, que representa un segundo caso de integración en nuestra praxis.

Integración metodológica

En la actualidad, a pesar de que el área de evaluación conductual se ha desarrollado ampliamente, se ha hecho necesaria la concurrencia de otro tipo de evaluación menos estructurada y más dirigida a tratar con los aspectos de validez de los temas a evaluarse. Nos referimos en este caso **a** la utilización de una metodología más de tipo cualitativa (ver Figura 16-4). Esta se hace necesaria sobre todo en las etapas iniciales de la investigación, cuando se considera un problema que, si bien puede existir una forma tipificada de evaluación para abordarlo, puede éste tener

características propias y diferentes a las que tendría ese mismo problema en otros contextos. Me refiero, por ejemplo, a aquellas ocasiones en las cuales podría pensarse que existen diferencias sociales y culturales que incidirían en un planteamiento diferente del mismo problema.
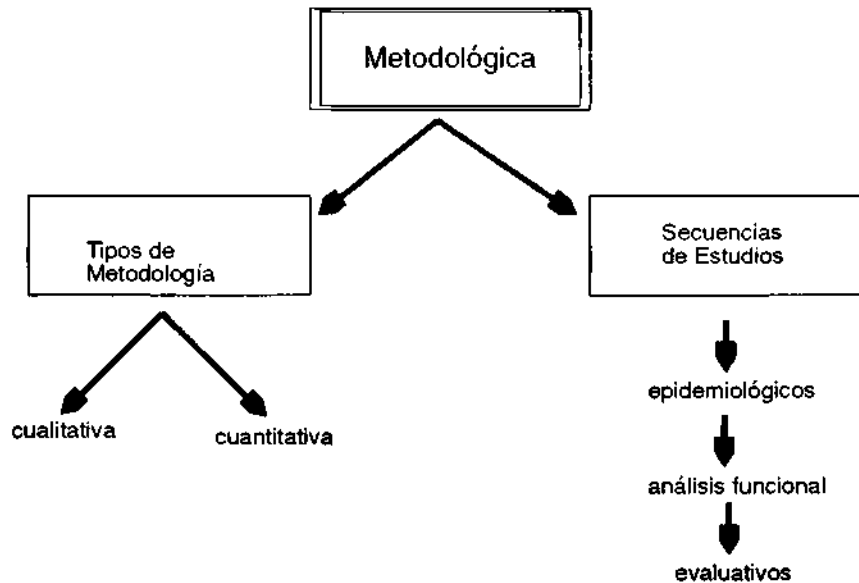
Fig. 16-4 — Nivel de integración metodologica.

Un ejemplo de ésta situación sería el caso de la evaluación de las destrezas o habilidades sociales. Si bien en la literatura se tipifica esta problemática de una manera más o menos clara, igualmente así su forma de evaluación, no podemos dar por sentado que las conductas concretas que representarían el constructo de habilidad social son las mismas en todas las culturas y subculturas. Por tanto se hace necesaria una exploración inicial de tipo cualitativa y observacional, la cual determina cuáles realmente son las conductas que representan, en un contexto dado, las habilidades sociales. Este tipo de trabajo describe el tercer tipo de integración, ya de otra naturaleza: la integración de la metodología cuantitativa con la metodología cualitativa. Si bien esta conjugación de métodos no es realmente una novedad, pues la tendencia actual conduce en esa dirección, también en la evaluación conductual que realizamos hemos detectado la necesidad de complementar una metodología positivista y cuantitativa con otra más dirigida al fenómeno mismo y a su observación sistemática, ésta con un propósito diferente a la determinación de frecuencias de ocurrencias.

Otro aspecto que caracteriza a la evaluación es la necesidad que nos planteamos de tener procedimientos e instrumentos que permitan la comparación tanto entre sujetos como entre grupos. Esto surge debido a que algunos de nuestros estudios son de **tipo epidemiológico,** lo cual hace necesaria la adopción de instrumentos evaluativos que permitan la realización de un análisis funcional lo más completo posible, a fin de no quedamos en las primeras etapas de determinación de índices de prevalência, sino también estar en capacidad de establecer relaciones entre variables (Vivas, Cañoto, 1990). Esta característica podría ilustrarse con el trabajo realizado sobre rechazo escolar, en el cual se creó una base de datos de tipo "epidemiológico" (Feldman, Granell, Vivas, 1985; Vivas, Granell, Gelfand, Feldman, 1982; Vivas, Feldman, Granell, 1990) tal como lo concebimos en el área de la Psicología (Granell, Vivas, Gelfand, Feldman, 1982, 1984; Granell, Feldman, Vivas, Gelfand, 1985, 1987), se determinaron las variables relacionadas con el problema estableciendo así índices de riesgo. Este tipo de secuencia de investigación, que abarca desde la base inicial — la epidemiologia — hasta la intervención, se ha hecho necesaria en nuestros países debido a la carencia de infraestructura o datos provenientes de investigaciones previas, y proporciona los elementos necesarios para, de una manera estable y confiable, elaborar programas preventivos más efectivos, que vayan dirigidos a aquellos grupos que efectivamente lo necesitarían.

Otro ejemplo en esta área de la investigación con fines comparativos es el trabajo que actualmente se está realizando en la elaboración de instrumentos, válidos para nuestro medio, que permitan confiablemente determinar tanto prevalência de factores psicosociales de riesgo y protectores de enfermedades cardiovasculares en nuestra población, como las relaciones entre ellos. En esta área, además de la validación de instrumentos específicos, se ha desarrollado todo un procedimiento de evaluación integral de los factores psicosociales asociados a enfermedades cardiovasculares (Bagés, Feldman, 1991; Feldman, Bagés, Chacón, Yanez, 1991). Dentro de este marco se sigue trabajando en instrumentos y procedimientos evaluativos en el área general de la salud.

Como puede observarse en los ejemplos antes mencionados, en este tipo de estudio predomina un interés más dirigido al trabajo con grupos *versus* el tratamiento de un sólo sujeto, si bien la evaluación se realiza a nivel individual. En este sentido nos separamos, de nuevo, de la orientación de investigación del sujeto unico, pero son las necesidades planteadas por la realidad de los objetos de estudio los que imponen el tipo de investigación a realizar, y no las preferencias del investigador por determinado modelo investigativo y conceptual.

Es de hacer notar, a modo de conclusión, que también en el aspecto metodológico se ha venido dando, planteado por la necesidad y la práctica,

otro tipo de integración. En esta oportunidad ya no se trata de integración de actividades, sino de integración de métodos y procedimientos que nos ayudan a evaluar, de la manera más completa y válida posible, la conducta en sus diferentes modalidades, asi como la interrelación de ésta con el medio.

Integración en la praxis

Una primera característica que determina el tipo de evaluación que realizamos es que generalmente ésta viene dada por las **necesidades** que se plantean en los **trabajos de investigación.** Dado que la investigación es una parte integrante fundamental de nuestra práctica docente, pues, como ya se mencionó al principio del capítulo, es la investigación la que fija la pauta docente y también de las actividades diarias de los mismos docentes, ésta es altamente tomada en cuenta. En ella se plantean problemas reales a solucionar y a los cuales necesariamente se impone dar respuesta. Por tanto se regresa muchas veces a la teoría y, por ende a la bibliografía, como queda ilustrado en la Figura 16-5. Este enriquecimiento constante se vierte directamente a los estudiantes, en la actividad docente. Se ha logrado una integración entre investigación y docencia. Acá podemos señalar como ejemplo que las necesidades metodológicas planteadas por los proyectos de investigación tanto de profesores como de estudiantes en proceso de elaborar su trabajo de grado proporcionan nuevos temas para la docencia, basados en experiencias directas y recientes.
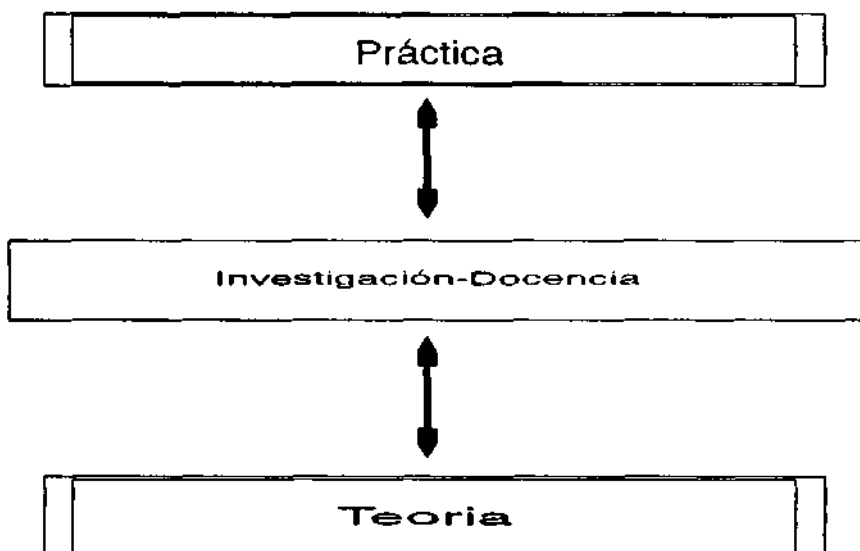
Fig. 16-5 — Nivel de integración práctica.

La consideración que podríamos plantear respecto de cual sería el desarrollo futuro de la evaluación conductual, partiendo del estado actual y proyectándose hacia el futuro, podría resumirse en los siguientes aspectos:

— La evaluación conductual, concebida fundamentalmente como una metodología, se va a relacionar cada vez más con el basamento conceptual que la sustenta. Portanto, su ambito de acción tenderá a ser más amplio cada vez. Esto puede evidenciarse en las actuales investigaciones, las cuales necesariamente evalúan tanto los tres canales de respuesta, como los tres elementos de la tríada que postula el determinismo recíproco. Se implica de esta manera que los basamentos teóricos deben ser de tipo comprehensivo, que expliquen de la manera más amplia y completa posible la conducta humana, de modo que los métodos para estudiarla puedan derivarse de estos de una manera fluida y que a la vez sean los más acertados.

— La metodología va a responder a las necesidades de los problemas que se abordan y, éstos a su vez, responden al objeto de estudio. Esto genera la necesidad de una metodología flexible, que se adapte a la diversidad de áreas y a la diversidad de componentes con sus correspondientes complejidades, a fin de conformar un cuadro lo más completo posible del individuo o grupo (ver Fig. 16-6).

— Hasta el momento se han evaluado conductas tanto abiertas como encubiertas y sus relaciones funcionales. Sin embargo, los **procesos** implicados han sido, si bien no ignorados, no abordados. Se impone un estudio más completo de estos procesos para conocerlos exhaustivamente y así estar en capacidad de evaluar la conducta en todas sus dimensiones. Este aspecto conlleva un nivel de análisis diferente, el cual va focalizado más bien hacia el "cómo" que hacia el "qué" de la conducta. Pero se hace necesario desarrollar una metodologia solida que permita estudiar los procesos de una manera valida y cientifica.

— Debido a que la meta u objetivo final de nuestro interés en el área de aplicación en la actualidad es fundamentalmente la promoción y prevención de la salud, la metodología utilizada está respondiendo cada vez más a esta meta. Hemos adoptado una metodología tanto epidemiológica como aquella que es necesaria para la implementación (Vivas, 1992) y evaluación de programas. Una vez realizada la evaluación de corte epidemiológico debería pasarse a investigaciones de tipo prospectivo para estudiar las relaciones de causalidad en los fenómenos que se abordan. Y finalmente, de allí, a la intervención para prevención y/o promoción.

— Si bien, de hecho, proporcionamos al estudiante herramientas para el abordaje de la evaluación de problemáticas individuales, el énfasis
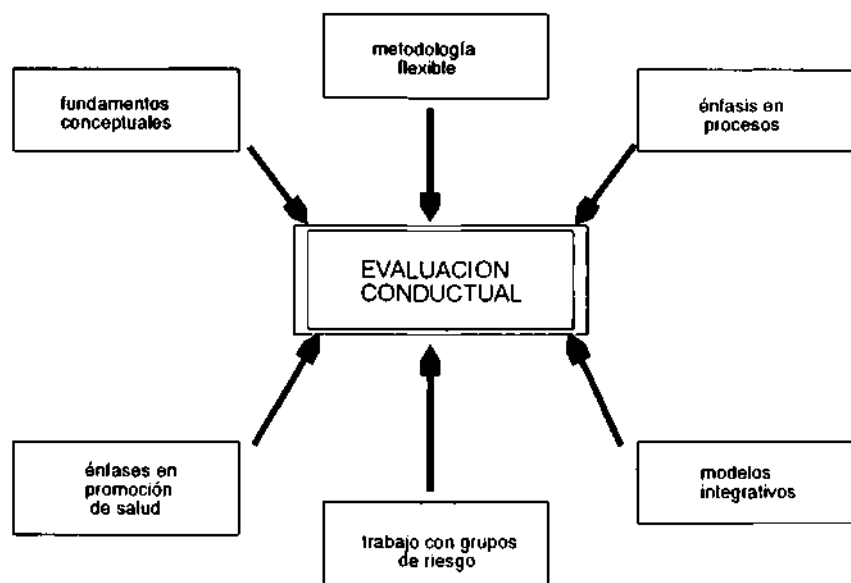
Fig. 16-6 — Perspectivas.

no se centra en éste, sino más bien en el trabajo con grupos de sujetos a fin de determinar factores que inciden en problemáticas de interés. Este aspecto se encuentra intimamente vinculado ai anterior, pues el interés preponderante hacia la promoción de salud nos lleva a contemplar, necesariamente, procedimientos tanto de evaluación como de intervención dirigidos a grupos y comunidades. Igualmente, esta práctica ha orientado algunas investigaciones hacia la simplificación de los instrumentos evaluadores, con el fin de que su uso pueda incorporarse en la práctica de otros profesionales afines, tales como médicos y miembros de la comunidad cuando ésto sea necesario o relevante. Esta fase, aunque incipiente, ha de sugerir algunas lineas de trabajo productivo en la evaluación conductual, que la vinculan con un modelo de intervención "desprofesionalizante" (Vivas, 1992) para atender a las cada vez más amplias necesidades de la realidad.

— Los métodos específicos tenderían cada vez más a estar amalgamados, sin hacer distinciones entre cualitativos y cuantitativos, pues dependen, como se dijo antes, del objeto de estudio y de la etapa en la cual se aborda el problema. Debido a la complementaridad que existe entre ambas metodologías, la integración metodológica es ya tanto una necesidad como un hecho que seguirá evidenciándose en la integración que se planteó en la praxis docencia-investigación.

— Como consecuencia de todos estos cambios integrativos y la amplitud que ya caracteriza a este tipo de evaluación, la cobertura de la misma se hace mucho más dilatada. Por ende, ya no se restringe su aplicabilidad a un área de la conducta ni su uso sólo por parte del profesional de orientación cognitivo-conductual. Debido a todos los aspectos que abarca en la actualidad, podría ser de gran utilidad para todos aquellos estudiosos de la Psicología que deseen lograr una evaluación integral del individuo. Y muy seguramente, en el futuro su uso será casi obligado para todos los profesionales de la Psicología.

Adicionalmente, desde el punto de vista académico y docente, es muy probable que sea incluida la evaluación cognitivo-conductual como una asignaura en los pensa de estudios, que represente una de las formas de recolección de información más completos sobre el sujeto y también sobre los grupos.

REFERENCIAS BIBLIOGRÁFICAS

BAGES, N., FELDMAN, L. *El medio laboral y las enfermedades cardiovasculares:* evaluación de los factores psicosociales de riesgo y protectores. San José, 1991. Trabajo presentado en el XXIII Congreso Interamericano de Psicología, Costa Rica, 1991.

BANDURA, A. *Social foundations of thought and action:* a social cognitive.. Englewood Cliffs (NJ): Prentice-Hall, 1986.

FELDMAN, L, BAGES, N.. CHACON, G., YANEZ, Y. *El estrés y los factores protectores en el medio organizacional.* Barcelona, 1991. Trabajo presentado en el Congreso Modificación de Conducta Latini Dies, España, 1991.

FELDMAN, L., GRANELL, E., VIVAS, E. Un estudio epidemiológico en niños venezolanos. III: Temor a los animales. *Psicología,* v.10, n.3/4, p.3-11, 1985.

FERNANDEZ BALLESTEROS, R. *El Ambiente:* análisis psicológico. Madrid: Pirámide, 1987.

-------------- . --------------- . 6.ed. Madrid: Pirámide, 1991.

GRANELL, E., VIVAS, E., GELFAND, D, FELDMAN, L Estimating the prevalence of school refusal and school related fears: a Venezuelan sample. *Journal of Nervous and Mental Disease,* Baltimore (MD), v.172, n.12, p.722-729,1984.

-------------- . Un estudio epidemiológico en niños venezolanos. I: Metodología. *Psicología,* v.9, n.4, p.653-665, 1982.

GRANELL, E., FELDMAN, L, VIVAS, E., GELFAND, D. Characteristics of Venezuelan school refusers: towards the development of a high-risk profile. *Journal of Nervous and Mental Disease,* Baltimore (MD), v.175, n.7, p.402-407, 1987.

-------------- . Desarrollo de un perfil de alto riesgo de rechazo escolar: algunos aspectos meto dológicos. *Boletín AVEPSO,* v.8, n.l, p.14-19,1985.

GRANELL, E., VIVAS, E., FELDMAN, L Dimensiones de problemas de conducta infantil según el maestro: análisis factorial en una muestra venezolana. *Interamerican Journal of Psychology,* v.22, n.1/2, p.75-96, 1988.

VIVAS, E. Análisis e ilustración de un modelo para la intervención en prevención. *Interamerican Journal of Psychology,* v.26, n.l, p.53-62,1992.

VIVAS, E., CAÑOTO, Y. *¿Estudios epidemiológicos en Psicología?* San José, 1990. Trabajo presentado en el XXIII Congreso Interamericano de Psicología, Costa Rica, 1990.

VIVAS, E., FELDMAN, L, GRANELL, E. Prevalência de problemas de lenguaje en niños venezolanos. *Comportamiento,* v.1, n.2, p.1-15,1990.

VIVAS, E., GRANELL, E., FELDMAN, L, GELFAND, D. Un estudio epidemiológico en niños venezolanos. II: Características de conducta. *Psicología,* v.10, n.1/2, p.3-25,1982.

ÌNDICE ONOMÁSTICO

INDICE DE ASSUNTOS